

Sistema de Ayuda al Diagnóstico para la Detección Temprana de Melanomas

Trabajo de Fin de Grado

Autor: José Serrano Fernández

26/09/2017



Tutor: Fernando Díaz de María

"A mi familia y amigos..."

Agradecimientos

En primer lugar quería darle las gracias a la Universidad Carlos III por permitirme desarrollar el grado que realmente se ajustaba a mis deseos. Gracias a los profesores que me han ayudado durante mi camino hacia el final del Grado en Ingeniería de Sistemas Audiovisuales.

Gracias a Fernando, cuyas clases en Tratamiento Digital de la Imagen me abrieron la mente hacia nuevos horizontes y me hizo perfilar el camino que actualmente quiero seguir en mi vida laboral, gracias además por orientarme en este trabajo y hacer que me pruebe a mí mismo y mejore mis aspiraciones.

Gracias a mis amigos por escucharme e interesarse por los interminables diálogos acerca de ingeniería que les doy, por comprender la dificultad de un grado de ingeniería y no presionarme en épocas de exámenes.

Gracias a mis compañeros de universidad por ayudarme y ser mi apoyo en tantas asignaturas y por ser un factor relevante en mi motivación para seguir creciendo como ingeniero.

Y gracias a mi familia, por ayudarme a levantarme tantas veces que he caído durante la carrera, por aguantarme en los momentos que la presión me hacía un poco insoportable, por valorar mis esfuerzos y no ponerme ningún tipo de traba para continuar.

En definitiva gracias.

Resumen

El cáncer de tipo melanoma tiene unas características que lo hacen diferente a otros cánceres de piel. Si el melanoma se encuentra en fases iniciales podrá extirparse de manera sencilla pero en fases más avanzadas, el melanoma es un tipo de cáncer que pone en alto riesgo la vida del paciente.

Sería interesante desarrollar un sistema que sirviera de ayuda para detectar este tipo de cáncer en fases tempranas y minimizar los riesgos del el melanoma.

El objetivo es crear un sistema de ayuda al diagnóstico para la detección in situ de melanomas y localización del cáncer de manera precoz.

El sistema de ayuda al diagnóstico precoz de melanomas es un sistema que emplea herramientas de tratamiento digital de la imagen y de aprendizaje automático y profundo. Con una base de datos de imágenes dermatoscópicas puede generar un modelo que clasifica estas imágenes como lesiones benignas o malignas, con la mayor exactitud posible.

Abstract

Melanoma has some characteristics that make it different from other skin cancers. If we diagnosed melanoma in the first phases of the illness, we would be able to remove it with minor surgery. On the other hand, in advanced phases of cancer, this kind of tumour can make connections through the blood to form metastasis, putting the patient's life at risk.

For these reasons, it will be interesting to develop a system to help us detecting this kind of cancer early, making it possible to easily reduce the risks associated with melanomas.

The objective is to create a computer-aided diagnosis system for *in-situ* detection of melanomas in early phases.

The proposed computer-aided diagnosis system for early melanoma detection is a system which relies on image processing, machine learning and deep learning. The system can generate a model which classifies the images of a data set either as benign lesions or melanomas.

Índice de Ilustraciones

Ilustración 1: Capas de la Epidermis.....	18
Ilustración 2: Ejemplo de cambio de tamaño de una lesión.	30
Ilustración 3: Ejemplo de cambio de forma de una lesión.....	30
Ilustración 4: Ejemplo de cambio de color de una lesión.	30
Ilustración 5: Ejemplo de diámetro no sospechoso y sospechoso de una lesión.	31
Ilustración 6: dermatoscopia tradicional.	36
Ilustración 7: dermatoscopia digital.	37
Ilustración 8: Ejemplos de imágenes dermatoscópicas	38
Ilustración 9: Cámara fotográfica utilizada en dermatoscopia digital	39
Ilustración 10: Teléfono Smartphone adaptado utilizado en dermatoscopia digital.....	39
Ilustración 11: Aparato dermatoscópico digital para el seguimiento de las lesiones.....	40
Ilustración 12: Esquema ilustrativo de modelo de aprendizaje automático utilizado.....	45
Ilustración 13: a) Imagen Original, b) Máscara, c) Imagen Segmentada.....	46
Ilustración 14: Matiz representado en 360 tonalidades sobre H.....	51
Ilustración 15: Redes neuronales organizadas en capas con una serie de nodos interconectados.	59
Ilustración 16: Diagrama de bloques de modelo en el que se incluye extracción de características por medio de CNN.....	59
Ilustración 17: Ejemplo de sección de matriz de datos médicos de las lesiones	66
Ilustración 18: Área bajo la curva de clasificador por medio de Árbol de Decisión.....	71
Ilustración 19: Área bajo la curva de clasificador por K-NN.....	72
Ilustración 20: Área bajo la curva de clasificador por SVM.....	74
Ilustración 21: Área bajo la curva del clasificador basado en Árbol de Decisión para la extracción de características por CNN	75
Ilustración 22: Área bajo la curva de clasificador por K-NN para la extracción de características por CNN	77
Ilustración 23: Área bajo la curva de clasificador por SVM en modelo de extracción de características por medio de CNN.....	78
Ilustración 24: Ejemplos de lesiones benignas clasificadas como melanomas.....	79
Ilustración 25: Ejemplos de melanomas clasificados como lesiones benignas.....	80

Índice de Tablas

Tabla 1: Estructura de matriz de concurrencia.	54
Tabla 2: Resultados para el clasificador por medio de Árbol de Decisión	71
Tabla 3: Matriz de confusión para el clasificador por medio de Árbol de Decisión.....	72
Tabla 4: Resultados para el clasificador por medio K-NN	73
Tabla 5: Matriz de confusión para el clasificador por K-NN.....	73
Tabla 6: Resultados para el clasificador por medio SVM	74
Tabla 7: Matriz de confusión para el clasificador por medio SVM.....	75
Tabla 8: Resultados para el clasificador por medio de Árbol de Decisión para la extracción de características por CNN	76
Tabla 9: Matriz de confusión para el clasificador por medio de Árbol de Decisión para la extracción de características por CNN	76
Tabla 10: Resultados para el clasificador por medio K-NN para la extracción de características por CNN	77
Tabla 11: Matriz de confusión para el clasificador por K-NN para la extracción de características por CNN	78
Tabla 12: Resultados para el clasificador por medio SVM en modelo de extracción de características por medio de CNN.....	79
Tabla 13: Matriz de confusión para el clasificador por medio SVM en modelo de extracción de características por medio de CNN.....	79
Tabla 14: Presupuesto de Recursos Físicos.	86
Tabla 15: Presupuesto de Recursos software	86
Tabla 16: Presupuesto de Recursos Humanos	87
Tabla 17: Presupuesto Total del Sistema se Ayuda al Diagnóstico Precoz de Melanomas	87

Índice General

Agradecimientos	3
Resumen.....	5
Abstract	6
Índice de Ilustraciones	7
Índice de Tablas.....	8
Índice General	9
1. Introducción y Objetivos	12
1.1. Introducción	12
1.2. Incidencia del cáncer de piel	14
1.3. Motivación	15
1.4. Objetivos	16
2. Estado del Arte	17
2.1. La Piel	17
2.1.1. La Epidermis	18
2.1.2. La Dermis.....	20
2.1.3. La Hipodermis.....	20
2.1.4. La piel como agente protector del organismo	20
2.2. El Cáncer de tipo Melanoma	21
2.2.1. Introducción al Melanoma: El cáncer de Piel	21
2.2.2. Causas.....	24
2.2.3. Factores de Riesgo.....	25
2.2.4. Detección Temprana	28
2.2.5. Signos y Síntomas	29
2.2.6. Diagnóstico	32
2.2.7. Clasificación por etapas.....	33
2.2.8. Práctica dermatológica.....	36
2.2.8.1. <i>Dermatoscopio tradicional</i>	36
2.2.8.2. Dermatoscopia Digital	37
2.3. Sistemas de Ayuda al Diagnóstico precoz de Melanomas	40
3. Diseño de sistema de ayuda al diagnóstico precoz de melanoma.....	43

3.1.	Visión General	43
3.2.	Descripción del Sistema	45
3.2.1.	Modelo de Aprendizaje Automático	45
3.2.1.1.	Extracción de Características diseñadas manualmente	48
3.2.1.1.1.	Descriptores basados en color	48
3.2.1.1.2.	Descriptores basados en forma.....	52
3.2.1.1.3.	Descriptores basados en Textura	53
3.2.1.1.4.	Selección óptima de características	57
3.2.1.1.5.	Normalización de las características	58
3.3.	Modelo con extracción de características mediante redes neuronales convolucionales.....	59
3.4.	Clasificación.....	61
3.4.1.	Clasificación por medio de árbol de Decisión	62
3.4.2.	Clasificación mediante K vecinos más próximos	62
3.4.3.	Clasificación mediante máquinas de vectores soporte (SVM).....	63
4.	Pruebas experimentales y resultados	65
4.1.	Descripción de la Base de datos.....	65
4.1.	Ensayos preliminares con descriptores diseñados manualmente	68
4.2.	Resultados	69
4.2.1.	Descriptores diseñados manualmente.....	70
4.2.2.	Descriptores obtenidos por medio de CNNs.....	75
4.2.3.	Errores cometidos	79
5.	Conclusiones y Líneas Futuras.....	82
5.1.	Conclusiones Obtenidas	82
5.2.	Líneas Futuras	83
6.	Marco Regulador.....	85
6.1.	Análisis de la legislación aplicable sobre la implementación descrita en el trabajo... 85	
7.	Entorno Socio Económico	86
7.1.	Presupuesto de la elaboración del TFG.....	86
8.	Referencias	88
	Anexo 1: English Summary	90
	Anexo 2: Referencias sobre el código utilizado	96

1. Introducción y Objetivos

1.1.Introducción

El cáncer de tipo melanoma muestra una serie de características que lo hacen único y diferente con respecto al resto de cánceres de piel. El melanoma es un cáncer que se desarrolla en los melanocitos, células que conforman los lunares, como consecuencia de ello se desarrolla en la propia piel.

Aquel melanoma que se encuentre en fases iniciales de desarrollo podrá extirparse de manera sencilla por medio de cirugía y sin necesidad de intervenciones quirúrgicas complicadas que ponen en riesgo la vida del paciente.

Como contrapartida, sin embargo, en fases más avanzadas, el cáncer de tipo melanoma establece fácilmente conexión con sistemas de comunicación presentes en el organismo, favoreciendo la aparición de metástasis o propagación de las células cancerígenas, a otras regiones corporales en las que puede encontrarse cualquier órgano vital. En este caso, el melanoma es un tipo de cáncer que pone en alto riesgo la vida del paciente.

Más adelante hablaremos de estadísticas referentes al desarrollo del cáncer de piel y concretamente de tipo melanoma, también haremos comentarios acerca de sus tasas de mortalidad en los últimos años en España y Estados Unidos, además de otros datos, pero en este momento comentaremos una serie de datos contrastados que ponen de manifiesto la relevancia que tiene la detección temprana del melanoma.

Ya hemos hablado de las facilidades que supone en cuanto a recuperación del paciente la detección temprana del melanoma, pero existen datos utilizados en la práctica clínica que reflejan la existencia de una delgada línea entre la posibilidad de extirpación y recuperación sencilla del cáncer en etapas tempranas, y la pasmosa dificultad de cura y alta mortalidad que puede suponer el melanoma en etapas muy avanzadas.

En la práctica médica, los especialistas suelen utilizar las tasas de supervivencia para valorar el pronóstico de una persona de una forma estándar. Para estudios en cáncer de tipo melanoma suelen emplearse tasas de supervivencia a 5 y 10 años, estas se refieren al porcentaje de los pacientes que viven por lo menos este tiempo después del diagnóstico de cáncer. Por supuesto, muchas personas viven mucho más de 5 o 10 años y muchas son curadas.

Las tasas de supervivencia [E1] se basan en los resultados previos de un gran número de personas que tuvieron la enfermedad; sin embargo, no pueden predecir lo que sucederá en el caso de cualquier persona.

Las diferencias en aspectos generales existentes entre las distintas etapas son las siguientes: un melanoma en etapa 1 es menor de 2mm, en etapa 2 es mayor de este grosor, en etapa 3 existe transmisión a los ganglios linfáticos y en etapa 4 además de todo lo anterior existe metástasis.

De manera resumida y yendo a los datos más drásticos podemos afirmar [E1]:

Para melanomas en etapa 1, en el peor de los casos con melanomas de hasta 2mm con ulceración la tasa de supervivencia de 10 años es del 86%

Para melanomas en etapa 2, en el peor de los casos con melanomas de más de 4mm con ulceración la tasa de supervivencia de 10 años es del 40%

Para melanomas en etapa 3, en el peor de los casos con melanomas de más de 4mm con ulceración y transmisión a los ganglios linfáticos la tasa de supervivencia de 10 años es del 24%

Para melanomas en etapa 4, en cualquier grosor cualquier tipo de transmisión a los ganglios linfáticos y con cualquier tipo de metástasis la tasa de supervivencia de 10 años es del 10% al 15%

Extrayendo conclusiones con estos datos podemos vislumbrar la fatídica realidad que representa detectar un melanoma en etapas avanzadas y la ventaja tan enorme que supone detectarlo en etapas tempranas. En este sentido este tipo de cáncer presenta características especiales, su detección a tiempo puede suponer la cura total con una cirugía sencilla y una pronta recuperación, pero en el momento en el que existe transmisión a los ganglios linfáticos y una posterior metástasis de células cancerígenas las posibilidades de recuperación se ven muy comprometidas, dándole mayor relevancia a la detección precoz al cáncer de tipo melanoma.

1.2. Incidencia del cáncer de piel

Después de una introducción acerca de las distintas tasas de recuperación en función de las etapas en las que se encuentra el melanoma, y que dejan ver la importancia que tiene la detección temprana de este tipo de cáncer, mencionaremos otra serie de datos que apoyarán esta idea.

Sabemos que además del melanoma existen distintos tipos de cáncer que se desarrollan en las células de la piel. Según datos recogidos en los últimos años, Podemos afirmar que el cáncer de piel es por mucho, el tipo de cáncer que más se desarrolla en las poblaciones. Analizando las estadísticas del cáncer de tipo melanoma, que forma parte de este conjunto, podemos decir, que sólo conforma el 1% entre todos los casos de cánceres de piel, sin embargo, es el que entrega mayor mortandad.

Según datos recogidos con anterioridad y cálculos de la Sociedad Americana Contra El Cáncer [E1], para el cáncer de tipo melanoma en los Estados Unidos en este año 2017, se estima que:

- Aproximadamente 87,110 nuevos casos de melanoma (alrededor de 52,170 en hombres y 34,940 en mujeres) serán diagnosticados.
- Aproximadamente 9,730 personas (6,380 hombres y 3,350 mujeres) morirán a causa de melanoma.

Los datos recogidos en EEUU reflejan que las tasas de aparición de melanoma han aumentado en los últimos 30 años. La frecuencia de aparición de melanoma es más de 20 veces mayor en las personas de raza blanca que en las de la raza negra. En general, el riesgo de padecer melanoma en el transcurso de la vida es de aproximadamente 2.4% para personas de raza blanca, 0.1% para personas de raza negra y 0.5% personas de raza amerindia. El riesgo para cada persona se puede afectar por un diferente número de factores, esta serie de factores los describiremos más adelante.

También se ha observado que el riesgo de padecer melanoma aumenta a medida que las personas envejecen. La edad promedio cuando se encuentra el melanoma es 62 años. Pero el melanoma es muy común incluso entre las personas menores de 30 años. De hecho, es uno de los cánceres más comunes en los adultos jóvenes, especialmente en mujeres jóvenes.

Según los datos obtenidos hasta Abril de 2016 y bajo la información brindada por Agustín Buendía, miembro de la Academia Española de Dermatología y Venereología, la situación en España es la siguiente [E4]:

Aumentan los casos de nuevos tipos de cáncer de piel, pero la tasa de mortalidad por melanoma se ha estabilizado.

La incidencia del cáncer de tipo melanoma en España es de 9,7 casos por cada 10.000 personas. La mortalidad causada por el mismo es de 2 personas por cada 100.000.

Para el resto de cánceres de piel la tasa de sitúa en 160 casos por 100.000 habitantes, de los cuales 118 son carcinomas baso celulares y 42 carcinomas espino celulares.

1.3.Motivación

Con todo lo mencionado con anterioridad y teniendo en cuenta la relevancia que tiene la detección y diagnóstico precoz de cualquier cáncer y en concreto el de tipo melanoma cutáneo, para poder tomar como opciones de tratamiento aquellas que garanticen una mayor calidad de vida y recuperación por parte del paciente, concluimos que sería interesante desarrollar un sistema que sirva de ayuda para detectar este tipo de cáncer en fases tempranas y favorecer la extracción sencilla y minimizar los riesgos del el melanoma.

Se sabe que, además, gran parte de los pacientes que acuden a los centros clínicos para hacerse un estudio dermatológico, no se han realizado previamente exámenes cutáneos y no han recibido ningún tipo de seguimiento en sus lesiones, ya sean o no pacientes con factores de riesgo; más adelante observaremos la relevancia que posee realizar este tipo de seguimientos. Por ello, desarrollar un sistema que ayude al diagnóstico in-situ de un posible desarrollo de cáncer de tipo melanoma, cuando no se ha recibido seguimiento de la lesión, puede resultar de gran ayuda en estos casos.

Generalmente los especialistas realizan una biopsia sobre la lesión cuando existe cualquier sospecha de melanoma. Este procedimiento genera una cicatriz en la piel del paciente, y sería interesante disponer de un sistema que ayudara a que se realizase únicamente cuando fuera estrictamente necesario.

El sistema en cuestión consistiría en un sistema de clasificación que, por medio de la imagen dermatoscópica de una lesión cutánea, detectase la probabilidad que tiene la lesión de ser una lesión benigna o haber desarrollado un melanoma.

También, este sistema podría servir como soporte para la docencia dirigida hacia dermatólogos con interés en dermatoscopia digital. Puede ser de ayuda para médicos no expertos en atención primaria que vayan a derivar al paciente al especialista posteriormente.

Finalmente concluimos que cualquier sistema de ayuda para la práctica médica para la detección temprana de melanomas sería de gran utilidad debido a las características de este tipo de cáncer y como afecta a la mortalidad la detección precoz del mismo.

1.4. Objetivos

El objetivo es desarrollar un sistema de ayuda al diagnóstico de cáncer de piel de tipo melanoma. Un sistema que favorezca y haga posible la detección temprana de este tipo de cáncer, creando un software a partir de la herramienta de cálculo Matlab con la que realizar un sistema de clasificación de imágenes dermatoscópicas que evalúe su pertenencia a uno de los dos siguientes grupos: lesión benigna y lesión maligna o melanoma.

Es un sistema de ayuda al diagnóstico que tiene como objetivo la detección in situ de melanomas. Servirá como herramienta de apoyo al especialista en casos en los que no se ha realizado seguimiento de las lesiones o lesión en cuestión, ni se dispone de otra serie de datos dermatológicos más que los de la propia imagen para poder contrastar datos. Esta imagen ha debido de ser tomada por medio de sistemas dermatoscópicos digitales normalizados bajo las mismas condiciones que las imágenes empleadas para el desarrollo inicial del sistema.

Será un sistema de apoyo útil en situaciones previas a una posible biopsia por sospecha de melanoma. Arrojará una segunda opinión al especialista acerca del desarrollo o no del cáncer bajo estudio, ya que el médico en cuestión, valorará además, otros factores clínicos del paciente: edad, antecedentes personales y familiares, tipo de piel, etc.

No tenemos intención de desarrollar un sistema de diagnóstico definitivo, con una fiabilidad absoluta, pues bajo los medios disponibles, únicamente imágenes dermatoscópicas y un sistema de software matemático, el análisis de las lesiones benignas y malignas resulta complicado debido la similitud que pueden presentar ciertas lesiones benignas con melanomas, así como la disparidad que existe entre las distintas lesiones benignas y la disparidad que también sufren entre sí las malignas.

2. Estado del Arte

2.1. La Piel

De manera introductoria, y para conocer más a fondo ciertas características del melanoma y el porqué de las mismas, haremos una breve descripción de la piel [E2][E3] como órgano corporal y saber de esta manera sobre qué células se desarrolla el cáncer de tipo melanoma entre otras cosas.

La piel es el órgano más grande del cuerpo humano. Posee un tamaño aproximado de 2 m^2 , constituyendo un 15-20% de la masa corporal total, con espesores que oscilan entre los 0.5 mm en los párpados, hasta los 4 mm en los talones, variando estas cantidades entre diferentes individuos.

La piel y sus derivados conforman el sistema tegumentario, el encargado de proteger al cuerpo frente a los agentes externos. Los derivados que forman parte de este sistema junto a la piel son: el cabello, las uñas, las glándulas sebáceas y las glándulas sudoríparas.

La función principal de este sistema tegumentario en nuestro organismo es la protección frente a factores externos como pueden ser: sustancias químicas, bacterias y otros microorganismos, temperatura y radiaciones. Más adelante hablaremos de la piel como agente protector del organismo.

La piel también contribuye a mantener la estructura del organismo, y se presenta como sistema comunicación del mismo con el entorno exterior a través de los distintos receptores que contiene, registrando temperatura, dolor, presión, tacto fino y grueso. Además, almacena agua y grasas, protegiéndonos de la deshidratación y permitiendo controlar la temperatura corporal. Otra función importante es la endocrina, secretando diversos factores de crecimiento y hormonas y permitiendo la formación de vitamina D.

Morfológicamente, la piel está compuesta por tres capas principales, ordenadas según su aparición de superficie a profundidad de la siguiente manera: epidermis, dermis e hipodermis. En la práctica diaria se tienen en cuenta principalmente las dos primeras.

Tras esta breve introducción comenzamos a describir con mayor profundidad las distintas capas de la piel que serán objeto de estudio.

2.1.1. La Epidermis

La epidermis, es la capa más externa de la piel, y se consolida como la barrera más importante que posee el cuerpo frente al entorno. En seres humanos su grosor varía entre 0.5 mm en párpados hasta 1.5 mm en la palma de las manos y planta de los pies. Es avascular, es decir, carece de riego sanguíneo, nutriéndose desde la dermis. Se encuentra constituida por dos grupos de células: queratinocitos o células no dendríticas y células dendríticas. Dentro de la epidermis, los queratinocitos, se organizan en 4 o 5 capas o estratos diferenciados.

Los queratinocitos, son las células encargadas de producir queratina y se constituyen como la célula más abundante dentro de la epidermis (95% de las células totales). Poseen estructuras de unión denominadas puentes intercelulares, que confieren estabilidad a la piel. Los queratinocitos cumplen con varias funciones. La más relevante es la de producir queratina, proporcionando resistencia al tejido.

Las capas o estratos que conforman la epidermis se ordenan de la siguiente manera desde el más externo hasta el más interno: estrato córneo, lúcido, granular, espinoso, y estrato basal. [E3] Se ilustran a continuación:

Estrato córneo: está formado por células que no poseen núcleo. Su grosor varía en función de la anatomía, siendo mayor en zonas como palmas y plantas.

Estrato lúcido: ubicado por debajo de la capa córnea, únicamente aparece en zonas de piel gruesa (palmas y plantas).

Estrato granuloso: formado por células romboidales que contienen gránulos de queratohialina, encargada de formar posteriormente los filamentos de queratina.

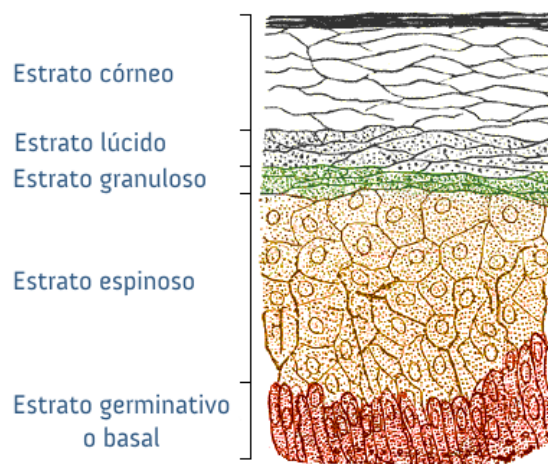


Ilustración 1: Capas de la Epidermis

Estrato espinoso: constituido por células poligonales que poseen unas prolongaciones conocidas como puentes intercelulares, estructuras que sirven como medio de unión entre las distintas células poligonales, confiriendo estabilidad al tejido. El número de estas células también varía dependiendo de la región corporal, en general de cinco a siete hileras.

Estrato basal, germinal o germinativo: está formado por células madre, de forma cilíndrica, que se disponen generalmente en una hilera. Tienen la capacidad de dividirse, formando las nuevas células cutáneas que renovarán el tejido periódicamente. Además, contienen la melanina sintetizada en los melanocitos.

La epidermis se mantiene gracias a la división continua de los queratinocitos en la capa basal. Estas células parten de la capa basal y se desplazan hasta las capas exteriores de la epidermis. Al llegar al estrato córneo, los queratinocitos pierden el núcleo y se mezclan con las capas escamosas que se desprenden continuamente de la superficie de la piel por un proceso conocido como descamación. Este recorrido tiene una duración aproximada de 4 semanas. En una piel sana la cantidad de células que son desprendidas es igual la que son producidas. La epidermis se renueva por completo en periodos de 48 días.

Como hemos mencionado las células no dendríticas o queratinocitos, no son el único tipo de célula perteneciente a la epidermis, ésta cuenta también con un segundo tipo son las células dendríticas, entre las que distinguimos: Melanocitos, células de Langerhans, y otras células. Las describimos a continuación:

Los melanocitos, llamados también células claras, son las células encargadas de generar la melanina, un pigmento. Podemos encontrarlas en el estrato basal, intercalados entre las células basales en una relación aproximada de un melanocito a diez células basales. Las proyecciones dendríticas de los melanocitos permiten el paso de melanina a los queratinocitos basales.

La melanina producida por estas células determina la coloración de la piel en el ser humano, en función de la cantidad y distribución de la misma, la piel se presentará más o menos oscura.

La melanina aparece almacenada en los melanosomas, partículas formadas en el interior de los melanocitos y que son transmitidas a los queratinocitos circundantes. El número, tamaño y organización de los melanosomas varía en los distintos grupos raciales: en pieles blancas u orientales los melanosomas se encuentran agrupados y en menor tamaño y cantidad que en las pieles oscuras o negras.

Al exponerse la piel a la radiación ultravioleta los melanocitos generan más melanina dentro de los melanosomas, haciendo que la piel se oscurezca. Así la melanina actúa como defensa de la piel frente a los rayos ultravioleta capaces de dañar el resto de células epiteliales.

Las células de Langerhans se originan en la médula ósea y se localizan en la piel y otras localizaciones como la mucosa oral, vaginal, ganglios linfáticos y timo. Dentro de la piel, se ubican principalmente en la epidermis y ocasionalmente en la dermis. La concentración de las células de Langerhans en la epidermis es semejante a la de los melanocitos: del 2 al 4% de la población epidérmica total. Una de las funciones principales de las células de Langerhans es la de fagocitar antígenos de microorganismos que pueden suponer un peligro para nuestra salud y presentárselos a las células del sistema inmune, previniendo de esta forma la infección.

2.1.2. La Dermis

La dermis es la capa de piel situada por debajo de la epidermis y está firmemente conectada a ella a través de la su membrana basal. Es más gruesa que la epidermis y está constituida principalmente por tejido conectivo, formado por tres tipos de fibras: fibras colágenas, elásticas y reticulares. Además, contiene folículos pilosos, glándulas sudoríparas, vasos sanguíneos y nervios que se mantienen en su sitio gracias a una proteína llamada colágeno, la cual imparte elasticidad y fuerza a la piel.

2.1.3. La Hipodermis

La hipodermis, o tejido celular subcutáneo, está constituido por células grasas, que se conocen con el nombre de adipocitos, los cuales se disponen en lóbulos separados por tejido conectivo.

Representa la capa más profunda de la piel y junto a la parte inferior de la dermis forman una red de colágeno y células adiposas. La hipodermis ayuda al cuerpo a conservar el calor y posee un efecto de amortiguación contra golpes que ayuda a proteger a los órganos internos.

2.1.4. La piel como agente protector del organismo

Ahora que conocemos las partes fundamentales de la piel es interesante comentar el papel que posee como agente protector del organismo.

La piel presenta diferentes alternativas a la hora de proteger el organismo contra agentes externos:

Como hemos visto, contiene células como las de Langerhans, capaces actuar frente a distintos microorganismos, activando al sistema inmune y evitando la propagación de la infección en el interior del organismo.

También sirve como regulador térmico ante los cambios de temperatura a los que el organismo se ve expuesto. Al entrar en contacto con temperaturas bajas los vasos sanguíneos de la dermis se contraen e impiden el paso de la sangre caliente proveniente del interior del organismo, además, la sangre que permanece en las zonas superficiales adquiere la temperatura del medio externo pero no puede viajar al interior, favoreciendo todo ello que se mantenga la temperatura corporal.

Por otro lado, los melanocitos sirven como defensa frente a los rayos ultravioleta capaces de dañar otras células del organismo.

Además, la piel recubre los órganos internos y ayuda a protegerlos de lesiones, evita la pérdida excesiva de agua y de otros líquidos, ayuda a controlar la temperatura corporal, protege el resto del cuerpo de los rayos ultravioleta (UV) y ayuda al cuerpo a producir vitamina D.

2.2.El Cáncer de tipo Melanoma

2.2.1. Introducción al Melanoma: El cáncer de Piel

Antes de entrar en materia acerca del cáncer de piel, haremos una breve descripción sobre qué es el cáncer, para poder comprender la magnitud y características del cáncer de tipo melanoma [E6].

Cada una de las células de nuestro organismo realiza ciertas funciones. Una célula sana se divide de manera ordenada y muere cuando se ha desgastado o dañado, y una célula nueva ocupa su lugar. El cáncer se origina cuando las células comienzan a crecer sin control, y acaban formando nuevas células que desplazan a las células sanas, generando problemas en el área del cuerpo en la que comenzó el cáncer. El cáncer, por tanto, puede originarse en cualquier parte del organismo.

Algunos tipos de cáncer crecen y se propagan rápidamente mientras que otros se multiplican más lentamente. Además, al ser diferentes tipos de células, actúan de formas muy diferentes, por lo que su sensibilidad a los tratamientos no es siempre la misma. Debido a ello, a menudo se administran dos o más tratamientos para obtener resultados óptimos. Existen diversos motivos por los que podría generarse un mismo tipo de cáncer, lo cual también dificulta el tratamiento y detección.

La mayoría de los cánceres forman una masa referida como tumor o crecimiento. Sin embargo, no todos los tumores son cancerosos o malignos. Las células cancerosas se pueden propagar a otras partes del cuerpo. La propagación de las células de cáncer a otras zonas del organismo se denomina *metástasis*. Aquellos tumores que pueden dar metástasis serán malignos, mientras que los que no tienen esa capacidad se conocen como tumores benignos.

En los casos en que un tumor maligno da metástasis, es difícil localizar el origen del cáncer, y como los distintos cánceres se tratan de maneras distintas dificulta en gran medida el tratamiento.

Es importante determinar que los cánceres se pueden clasificar en etapas en función del grado de propagación del mismo, partiendo de la etapa 1 como fases iniciales del cáncer hasta la etapa 4 que significa una mayor propagación o aparición de metástasis.

El cáncer de piel es un tipo de cáncer originado en las células de la piel. Existirán por tanto distintos tipos de cáncer de piel en función de en qué tipo de célula se origine.

Los tipos de cáncer de piel más conocidos son los siguientes [E1]: Carcinoma de células basales, Carcinoma de células escamosas y Melanoma. Existen más tipos de cáncer que se originan en la piel, como el Carcinoma de las células de Merkel, pero todos ellos presentan menos del 1% de los cánceres de piel registrados por lo que no haremos hincapié en ellos.

Ha de tenerse en cuenta que este tipo de cáncer puede darse en cualquier parte del cuerpo, pero su aparición es más común en zonas que son expuestas con asiduidad a las radiaciones solares como la zona de la cara, el cuello, las manos y los brazos.

Realizaremos una breve descripción de los distintos cánceres de piel para poder diferenciarlos del melanoma.

2.2.1.1. Carcinoma de células Basales

El carcinoma de células basales es el tipo más común de cáncer de piel, en torno al 80% de los casos de cáncer de piel. Se origina en la capa basal de la epidermis.

Por lo general, estos cánceres surgen en las zonas expuestas al sol, especialmente la cabeza y el cuello. Son tumores de crecimiento lento y es muy poco común que se propague a otras partes del cuerpo. No obstante, si no se recibe tratamiento adecuado, puede crecer localmente, destruyendo los tejidos adyacentes tales como hueso, músculo o los órganos situados bajo él.

Si no se extrae completamente, el carcinoma de células basales puede reaparecer en el mismo lugar que se originó. Las personas que han tenido cánceres de piel de células basales también tienen una probabilidad mayor de padecer nuevos cánceres en otros lugares, principalmente por el estímulo de radiación ultravioleta recibido a lo largo de los años. Si no se extrae completamente, el carcinoma de células basales puede reaparecer en el mismo lugar de la piel. Las personas que han tenido cánceres de piel de células basales también tienen una probabilidad mayor de padecer nuevos cánceres en otros lugares.

2.2.1.2. Carcinoma de células Escamosas

Alrededor del 20% de los casos de cáncer de piel son carcinomas de células escamosas. La morfología de las células que los conforman recuerda a versiones anormales de las células escamosas vistas en las capas externas de la piel.

Estos cánceres aparecen comúnmente, al igual que el de células basales, en áreas del cuerpo expuestas al sol, tales como la cara, las orejas, los labios y el dorso de las manos. También puede surgir en cicatrices o llagas crónicas de la piel en otras partes del cuerpo. Con menor frecuencia, se forman en la piel del área genital.

Los cánceres de células escamosas tienen más probabilidad de crecer hacia las capas más profundas de la piel y propagarse a otras partes del cuerpo que los cánceres de células basales, aunque esto sigue siendo poco común.

2.2.1.3. Melanoma

El cáncer producido a partir de los melanocitos se denomina melanomas [E7] [E13]. Este tipo de cáncer también es denominado melanoma maligno o melanoma cutáneo. Sabemos que los melanocitos también pueden generar crecimientos benignos o lunares.

Dentro de los cánceres de piel, el de tipo melanoma es el menos común. Sin embargo, ha de tenerse muy en cuenta ya que la proporción de mortalidad que presenta el melanoma con respecto al resto de cánceres de piel es muy superior. Tiene un crecimiento mayor, y mucha mayor facilidad para propagarse y dar metástasis si no se trata a tiempo. Al igual que el resto de cánceres de piel, es casi siempre curable en sus etapas iniciales, por lo que cobra una gran importancia su diagnóstico precoz.

Una vez desarrollado el cáncer, en muchas ocasiones, las células del melanoma siguen produciendo melanina de modo que suelen presentar tonos café o negro. Sin embargo existen ocasiones en las que esto no es así por lo que también presentan tonos rosados o incluso blancos.

Los melanomas pueden desarrollarse en cualquier parte de la piel, pero generalmente son más propensos a aparecer en el tronco, tanto pecho y espalda en los hombres y en las piernas en las mujeres. Otras zonas con alta exposición a los rayos solares como el rostro o el cuello son también sitios muy comunes. Las personas con pieles oscuras son menos propensas a padecer el melanoma pero tiene posibilidades de desarrollarlo, sobre todo en palmas de las manos o plantas de los pies y en ellos es más peligroso por

la dificultad de su diagnóstico. En ocasiones el melanoma se produce incluso en ojos, boca, genitales y área anal, pero son zonas mucho menos comunes.

Durante los siguientes apartados hablaremos de los distintos síntomas, sistemas de prevención, factores de riesgo y etapas entre otros, acerca del melanoma, y reflejaremos de nuevo, la relevancia que posee realizar un diagnóstico preciso y precoz para lograr superar con éxito esta enfermedad.

2.2.2. Causas

Existen una serie de factores que los investigadores han descubierto que pueden aumentar las posibilidades de que una persona padezca melanoma pero no son capaces de concretar exactamente como estos factores causan directamente la enfermedad.

Se han hallado una serie de cambios genéticos dentro de las células de lunares benignos que pueden causar que estas se transformen en células de melanoma pero no se sabe porque los lunares se tornan cancerosos.

Tras realizar varios estudios se ha demostrado que puede que el cáncer sea causado por cambios en el ADN. Estos cambios provocan que se activen los genes cancerígenos y se desactiven los genes supresores de tumores que se encuentran dentro de la cadena. Para ello sería necesario que ocurrieran cambios en varios genes distintos para que una única célula se tornara cancerosa.

Algo que se afirma con mayor seguridad es el papel tan relevante que juegan los rayos ultravioleta como causa principal de melanomas, ya que pueden dañar el ADN de las células de la piel y afectar a genes que controlan como crecen y se dividen pudiendo provocarse el cáncer. A pesar de ello también existen cánceres que se desarrollan en zonas cuya exposición a los rayos ultravioleta ha sido muy reducida o prácticamente nula por lo que no podemos otorgarle a estos toda la responsabilidad, más cuando estos melanomas se desarrollan de distinta manera que los que aparecen en zonas con mucha exposición.

Como hemos comentado se sabe de los riesgos que presentan los diferentes factores en la piel pero no se sabe la causa exacta de porque se genera el cáncer a partir de ellos ni mucho menos en todos los casos, es por ello que toda medida preventiva es poca ya que no disponemos de la causa directa del suceso ni cómo abordarla.

2.2.3. Factores de Riesgo

Un factor de riesgo es todo aquello que puede afectar a que un individuo padezca una enfermedad, en este caso un cáncer de tipo melanoma. Existen ciertos factores de riesgo que pueden controlarse, como la exposición excesiva a las radiaciones solares, pero existen otros factores que escapan al control del paciente, como los antecedentes familiares. A pesar de ello ha de tenerse en cuenta que el hecho de que un individuo posea uno o varios factores de riesgo no implica que vaya a padecer necesariamente cáncer de tipo melanoma, de la misma manera, no poseer ningún factor de riesgo o poseer pocos, no le exime de la posibilidad de padecer el cáncer.

Sin embargo, existen varios factores de riesgo que pueden afectar a que el individuo tenga mayores posibilidades de padecer un melanoma.

Exposición a la luz ultravioleta (UV): La exposición a los rayos ultravioleta es uno de los principales factores de riesgo para la aparición de la mayoría de los melanomas. Las principales fuentes de luz ultravioleta son, la luz solar, aunque los rayos UV solo representan una pequeña porción de la misma (solo determinadas longitudes de onda), y las camas bronceadoras o lámparas solares, donde los efectos de estas últimas suelen ser completamente evitables.

Los rayos UV dañan el ADN de las células de la piel, es decir son los principales causantes del deterioro de la piel, al formar parte de las radiaciones solares es imprescindible acotar sus riesgos. Cuando el daño producido por estas radiaciones afecta al ADN de los genes que controlan el crecimiento de las células de la piel se origina el cáncer de piel, por lo que la exposición a la misma podría fomentar el desarrollo de un melanoma en la piel del individuo

Los expertos consideran que los cánceres que se originan en zonas de la piel que se ven expuestas a las radiaciones solares de manera ocasional como pueden ser zonas como el tronco o las piernas y que han sido motivados por quemaduras en los primeros años de vida son diferentes a los originados en zonas acostumbradas a la exposición a las radiaciones solares.

Lunares: Generalmente los lunares empiezan a aparecer en la infancia o al llegar a la juventud adulta, es decir, no son innatos. Sabemos que la mayoría de los lunares no causaran ningún problema, pero en caso de que el individuo posea muchos, será más propenso a padecer cáncer de tipo melanoma.

A continuación describiremos los principales tipos de lunares y el riesgo que presenta cada uno de ellos a la hora de presentar un papel en el desarrollo del cáncer.

- **Lunares atípicos o nevos displásicos:** Este tipo de lunar tienen similitudes tanto con los lunares normales pero presentan características pertenecientes al melanoma. Suelen ser más grandes y tener formas y colores anormales, pueden aparecer tanto en zonas con alta exposición a los rayos solares como las cubiertas. generalmente son hereditarios, y la mayoría nunca se vuelven

cancerosos, de hecho muchos melanomas se originan sin este tipo de lunar, a pesar de ello, existe un pequeño riesgo de convertirse en melanomas

- **Síndrome de nevos displásicos:** también conocido como síndrome del melanoma de nevo múltiple atípico familiar, o FAMMM, es un síndrome de carácter hereditario. Las personas que lo padecen poseen gran cantidad de nevos displásicos y al menos un pariente cercano que ha tenido un cáncer de tipo melanoma, lo que genera muy riesgo muy alto de que la persona desarrolle este tipo de cáncer en algún momento de su vida. Teniendo en cuenta el riesgo que supone esta afección, esta serie de personas necesitan exámenes regulares y minuciosos de su piel por parte de médicos especialistas de la piel.

Existen cada vez más mayores herramientas que permiten hacer registros temporales de la piel de los individuos, no solo para la práctica médica, si no para la realización de autoexámenes, que en ocasiones juegan un papel importante a la hora de localizar a tiempo cualquier tipo de cáncer de piel. Más adelante hablaremos de este tipo de herramientas así como de otro tipo de máquinas empleadas en el tratamiento y la prevención del cáncer de piel.

- **Nevos melanocíticos congénitos:** En contraposición a los nevos displásicos existen este tipo de nevo melanocítico que está presente en el nacimiento del individuo. Existe en el caso de estos lunares una posibilidad de aparición de melanoma a lo largo de la vida del paciente que se estima entre el 0 y 10 %, dependiendo de tamaño del nevo, a mayor tamaño las probabilidades aumentan. Existe además un mayor riesgo según la zona de la piel en la que estén, en zonas como la espalda o los glúteos las posibilidades de aparición del cáncer serán significativamente más elevadas que en caso de estar en la palma de la mano.

Son, por tanto, el tamaño, la zona aparición y color del nevo, razones a tener en cuenta con respecto a la prevención, sin embargo, en muchos casos, como medida de seguridad, los médicos prefieren extirpar quirúrgicamente este tipo de nevos. En caso de no ser extirpados se recomienda realizar exámenes periódicos por parte de un dermatólogo así como autoexámenes como en casos anteriores. La probabilidad de aparición de melanoma sigue siendo baja, pero debido a la existencia de posibilidades se recomienda prevención, sobre todo en los casos de mayor riesgo anteriormente mencionados.

Piel muy blanca, pecas y cabello claro: Como ya hemos comentado los individuos con fenotipo tipo 1, caucásicos poseen mayores posibilidades de padecer el cáncer, siendo aquellos que además posean cabello y ojos claros los grupos de mayor riesgo

Antecedente familiar de melanoma. Su riesgo de padecer un melanoma es mayor si uno o varios de sus familiares de primer grado han tenido melanoma. Aproximadamente 10% de todas las personas con melanoma tienen antecedentes familiares de esta enfermedad.

El aumento en el riesgo pudiera deberse a que compartieron un estilo de vida de exposición frecuente al sol en la familia, una familia de piel muy blanca, ciertos cambios genéticos que son más frecuentes en una familia, o una combinación de factores.

Los expertos recomiendan que estas personas realicen una serie de procedimientos habituales o rutinarios, con el fin de prevenir la aparición o detectar a tiempo el cáncer.

Antecedente personal de melanoma u otros cánceres de piel: Una persona que ha padecido melanoma tiene un mayor riesgo de padecer melanoma nuevamente. Alrededor del 5% de las personas con melanoma padecerán un segundo melanoma en algún momento de sus vidas. Las personas que han tenido cáncer de piel de células basales o escamosas también están en mayor riesgo de padecer melanoma.

Sistema inmunológico debilitado: El sistema inmunológico ayuda a proteger el cuerpo del individuo y por tanto a combatir el cáncer. Las personas cuyos sistemas inmunológicos se han debilitado, debido a ciertas enfermedades, tratamientos médicos, etc., tienen más probabilidades de padecer cáncer de piel, éste puede ser de varios tipos, incluyendo melanoma.

Razones como: recibir un órgano trasplantado, en cuyo procedimiento se administran medicamentos que debilita el sistema inmunológico para evitar el rechazo del cuerpo hacia el nuevo órgano; así como personas infectadas con el VIH cuyo sistema inmunológico también está debilitado, son razones de riesgo de padecimiento de melanoma.

Una edad avanzada: Existe mayor probabilidad de aparición del melanoma en edades avanzadas, sin embargo, el cáncer de tipo melanoma, se detecta más que otros tipos de cáncer en personas de menos de 30 años, especialmente mujeres, es decir, es de los cánceres que más se diagnostican en personas jóvenes. Los melanomas que se presentan con antecedentes familiares pueden darse también en edades tempranas.

Sexo masculino: Según estadísticas, en EEUU, el sexo masculino posee una tasa mayor de melanoma que el femenino, sin embargo, esto varía con la edad. Para edades previas a los 45 años las mujeres poseen un mayor riesgo de padecimiento, mientras que posteriormente el riesgo es superior para los hombres. Son datos recogidos únicamente en una región del planeta pero pueden extrapolarse conclusiones con cierto grado de confianza.

Xeroderma pigmentoso: El Xeroderma pigmentoso o XP afecta a la capacidad que poseen las células de la piel para repararse tras haber sufrido un daño en su ADN. Este padecimiento es de carácter hereditario y las personas que lo poseen tienen un alto riesgo de padecer, en edades tempranas, uno entre los distintos tipos de cánceres de piel, entre los que destaca el melanoma por su letalidad. Este riesgo es más alto sobre todo en áreas que han sido expuestas a las radiaciones solares.

2.2.4. Detección Temprana

Autoexamen de la piel: [E5] Consisten en hacer una revisión de nuestra piel, para poder conocer nuestros lunares, patrones de los mismos, pecas marcas e imperfecciones con el fin de tener la capacidad de detectar si existe un cambio en la superficie de la piel, que aparezca un nuevo lunar o uno de los existentes sufra un cambio de tamaño, color, o cualquier cosa que nos llame la atención y consideremos que no es usual.

Lo ideal es realizar un autoexamen una vez al mes, normalmente se realiza en una habitación bien iluminada y frente a un espejo es cuerpo entero empleando otro de menos tamaño para poder ojear áreas de la piel con un acceso más difícil. Debemos poder hacer un registro de toda la superficie de la piel y solicitar ayuda en caso de que la necesitemos para zonas como la espalda ya que es aquí donde más melanomas se registran a lo largo del año en hombres. No debe olvidarse el cuero cabelludo y las uñas.

Hoy en día contamos con tecnología al alcance de nuestra mano para poder tener una ayuda a la hora de realizar el registro del mapa de nuestra piel por medio de una serie de tecnologías. Éstas pueden sernos de gran ayuda para realizar un estudio más eficaz. Las introduciremos más a fondo más adelante.

Examen por un profesional de la salud: Un profesional de la salud o medico realizara un examen minucioso de su piel como parte de un examen de rutina relacionado con el cáncer, porque posee un factor de riesgo, o sospechas de que un lunar ha cambiado o ha aparecido. Le asistirá en cualquier resultado que dé el examen así como cualquier información que necesite..

Para poder realizar esta serie de exámenes se emplean una serie de procedimientos que comentamos a continuación:

Normalmente el médico de cabecera descubre cualquier lunar inusual u otras áreas que causan sospechas, y recomienda al paciente asistir al dermatólogo o médico especializado en problemas de la piel. Los dermatólogos pueden hacer exámenes regulares de la piel.

El primer proceso a la hora de determinar si existe o no melanoma consiste en emplear una técnica llamada dermatoscopia o microscopia de epiluminiscencia para poder observar la piel con más claridad, además se puede tomar una fotografía digital de la imagen. El proceso completo de esta técnica lo desarrollaremos más adelante.

Posteriormente si existe sospecha de que es un melanoma normalmente se procede a realizar una biopsia, la cual se realizará de una forma u otra en función del estado en el que se encuentre la zona de sospecha según el criterio del médico. Más adelante lo desarrollaremos el proceso de diagnóstico de melanoma con mayor profundidad. Los exámenes regulares de la piel son especialmente importantes para las personas que presentan un mayor riesgo de melanoma, es decir, personas que presentan factores de

riesgo tales como el síndrome del nevo displásico, un fuerte antecedente familiar de melanoma, y padecimiento de melanoma en el pasado.

Para aquellas personas que presentan los factores de riesgo mencionados, o personas que poseen gran cantidad de lunares, entre otras, será interesante que su médico recomendara tomar fotos de todo el cuerpo para realizar un registro de los mismos, de modo que los lunares puedan ser rastreados con el pasar del tiempo y en caso de aparecer alguno nuevos puedan ser detectados, así como cualquier tipo de anomalía. Esta técnica se le llama fotografía corporal total o mapeo de lunares, existen una serie de tecnologías que también comentaremos más adelante.

2.2.5. Signos y Síntomas

Sabemos que un lunar normal morfológicamente se presenta como una mancha uniforme en la piel de color café, canela o negro y puede ser plano o prominente, redondo u ovalado. Además generalmente, los lunares miden menos de 6 milímetros de ancho, aproximadamente el ancho de una goma de lápiz.

Teniendo además en cuenta la descripción de las distintas clases de lunares, sobre las que puede desarrollarse cáncer, que hemos visto en apartados anteriores, podemos describir la regla que se emplea por antonomasia para poder identificar las señales habituales del melanoma, La regla o criterio **ABCD**.

2.2.5.1. *Criterio ABCD*

Este criterio o regla consiste en las siguientes características:

A de Asimetría: la mitad del lunar o marca de nacimiento no corresponde a la otra mitad.

B de Borde: los bordes son irregulares, desiguales, dentados o poco definidos.

C de Color: el color no es uniforme y pudiera incluir sombras color marrón o negras, o algunas veces con manchas rosadas, rojas, azules o blancas.

D de Diámetro: el lunar mide más de 6 milímetros de ancho (aproximadamente del tamaño del borrador de un lápiz), esto no quiere decir que los melanomas no puedan ser más pequeños que esto.

E de Evolución: el tamaño, la forma o el color del lunar están cambiando a lo largo del tiempo.

Es necesario asistir a un médico si se encuentra alguna de esas características en el momento en el que se realiza un autoexamen completo, o simplemente lo hemos detectado de manera esporádica.

Es una técnica cuyo nombre es fácil de aprender y muy interesante a la hora de realizar un autoexamen ya que estas son los principales síntomas visuales que presenta un melanoma.

2.2.5.2. Criterio de 7 puntos de Glasgow

La lista de control de 7 puntos de Glasgow [E9] consiste en una serie de criterios algunos de los cuales difieren en relevancia y se emplean para poder determinar la posibilidad que tiene un lunar o lesión de la piel de presentar cáncer de tipo melanoma. Los tres primeros criterios son los que poseen mayor importancia y habrá que considerarlos por delante de los siguientes cuatro criterios.

Criterios de mayor relevancia o criterios mayores:

- **Cambio de tamaño:**



Ilustración 2: Ejemplo de cambio de tamaño de una lesión.

- **Cambio de forma:**



Ilustración 3: Ejemplo de cambio de forma de una lesión.

- **Cambio de color:**



Ilustración 4: Ejemplo de cambio de color de una lesión.

Criterios menores:

- **Diámetro:** superior o igual a 7 mm



Ilustración 5: Ejemplo de diámetro no sospechoso y sospechoso de una lesión.

- **Inflamación**
- **Sangrado**
- **Cambio de sensibilidad:** Se siente picor o dolor en la propia lesión.

A la hora de ponderar los criterios anteriormente mencionados habrá que tenerse en cuenta que los criterios mayores equivaldrán a 2 puntos cada uno de ellos y los criterios menores equivaldrán únicamente a 1 punto.

Existirá sospecha de desarrollo de cáncer de piel de tipo melanoma cuando una lesión presente, por medio del empleo de los criterios mencionados, una suma de 3 o más puntos según las ponderaciones descritas. Es decir, dos criterios mayores provocaran alarma y tres menores o uno mayor y uno menor también, por ejemplo.

Como hemos comprobado, los criterios de Glasgow hacen un mayor hincapié sobre los cambios sufridos por la lesión y no tanto sobre su aspecto in-situ, esto permite detectar melanomas no considerados por la regla ABCD, por ello, es interesante utilizar ambos criterios.

Ambos criterios son válidos y es recomendable emplearlos conjuntamente para rechazar toda posibilidad. Ha de tenerse en cuenta que cada persona, al margen de la cantidad de lesiones o lunares que posean, presentan todos ellos aspectos similares, o cierto parecido, por lo que por lo general emplear el signo denominado del “patito feo”, es decir, considerar aquel que difiera demasiado de los demás como sospechoso, será de gran utilidad como complemento a los criterios anteriormente mencionados.

El dolor en uno de los lunares es también un motivo de sospecha de aparición de melanoma, sin embargo, la mayoría de los cánceres de este tipo carecen de este tipo de síntoma por lo que en caso de no aparecer no debe de ser descartado como maligno.

Aparte de lo mencionado, aspectos presentes en los lunares como: un brillo inusual, una superficie áspera, picor o que parezcan una herida sin cicatrizar, son también motivos de sospecha.

Es interesante tener en cuenta todos los factores [E10] mencionados para poder realizar un diagnóstico con exactitud, en caso de sospecha de melanoma se procederá a evaluar el lunar en cuestión por medio de procesos que comentaremos a continuación, para poder descartar o no la aparición o desarrollo del cáncer en el mismo.

2.2.6. Diagnóstico

Posteriormente al estudio de la piel del paciente, con los métodos mencionados y teniendo en cuenta los factores de riesgo que posee el paciente, extraerá la conclusión de si existe o no sospecha en alguna de las lesiones del paciente.

En caso de sospecha, por parte del médico, de que pueda haberse desarrollado melanoma sobre una lesión, procederá a tomar una muestra de la misma y la enviará a un laboratorio para que pueda ser examinada por un microscopio este proceso se denomina biopsia de la piel.

Existen muchos tipos de biopsia de la piel. El especialista procederá a realizar un tipo u otro en función de una serie de factores: la localización en el cuerpo, el tamaño del área afectado entre otros factores.

El proceso de biopsia se realiza por medio de anestesia local, debido a ello es indoloro, sin embargo, es propenso a dejar cicatriz, ésta será distinta y tendrá distintos tamaños en función del tipo de biopsia realizada. Ha de tenerse en cuenta que este es el menor de los problemas con tal de realizar el diagnóstico más preciso posible por lo que será preferible extraer tanta área sospechosa como sea posible.

Se sabe que en ocasiones puede ser necesario realizar biopsias en otros tejidos que no sean la piel. Si se diagnostica melanoma se suele realizar biopsias de ganglios linfáticos adyacentes a la lesión para determinar si el cáncer se ha propagado a ellos.

En ocasiones los melanomas se propagan con tanta rapidez que alcanzan órganos internos y otras zonas mientras que el melanoma original sigue siendo muy pequeño, es difícil que se haya hecho una biopsia ya que su detección es complicada, En ocasiones se descubren incluso cuando el melanoma ya se ha extirpado.

También sucede que para cuando se detecta el cáncer en otro lugar del cuerpo de la piel, la lesión con la que se originó haya desaparecido ya que en ocasiones desaparecen por si solas, pero algunas de estas células ya se ha propagado por el resto del cuerpo, determinar el origen en estos casos suele ser complicado

Es en estos casos de propagación del melanoma cuando hay que tener extremo cuidado porque puede confundirse con un cáncer de otro órgano implicado en la propagación, es decir, que tiene origen en ese órgano, y no en la piel como realmente

es. En estas ocasiones cuando la biopsia juega un papel relevante ya que se emplea cual es el tipo de cáncer, y en función de los resultados se procederá a un tipo de tratamiento u otro en función del tipo de cáncer que se determine.

Las biopsias en zonas internas del organismo son más complejas que las superficiales mencionadas anteriormente

Una vez obtenidas las muestras por medio de biopsia, son enviadas a un laboratorio donde un doctor patólogo normalmente un dermatólogo las observará por medio de un microscopio para determinar si hay células de melanoma.

Si el especialista determina que las muestras contienen melanomas, buscará características relevantes para poder catalogar el melanoma en una etapa o estadio como los anteriormente mencionados, y así poder elegir de manera efectiva entre las distintas opciones de tratamiento, ya que el pronóstico se verá afectado.

2.2.7. Clasificación por etapas

La etapa o estadio de un cáncer es una descripción de cuánto se ha propagado la enfermedad. Para el melanoma la etapa en la que se presente se determinará en función de factores como su grosor en la piel, si ha existido propagación a los ganglios linfáticos o a cualquier otro órgano entre otros. La determinación de en qué etapa se sitúa el cáncer se realiza por medio de resultado de exámenes físicos, biopsias u otros estudios realizados.

La etapa [E11] en la que se presenta el melanoma es muy importante para planificar su tratamiento y hacer un su pronóstico aproximado.

Existen dos formas de clasificar por etapas un melanoma:

La **clasificación clínica por etapas** se realiza en base a lo que se detecta en los exámenes físicos, biopsia o extirpación del melanoma principal y cualquier estudio que se lleve a cabo.

La **clasificación patológica por etapas** utiliza toda la información anterior y además lo que se haya detectado en las biopsias de los ganglios linfáticos o de otros órganos, si es que se realizaron.

Los médicos prefieren usar la etapa patológica ya que ofrece una idea más real y precisa de la extensión y magnitud del cáncer, simplemente en ocasiones no son necesarias las biopsias que permiten utilizar este tipo de clasificación.

El sistema más empleado en la práctica médica para clasificar el cáncer de tipo melanoma por etapas es el sistema TNM, formulado por la American Joint Commission on Cancer (AJCC), y se basa en tres piezas clave:

La letra **T** significa que se ha producido un tumor, se le asigna un número del 0 al 4 que informa acerca del grosor del mismo, es decir, la profundidad con la que ha crecido. Se asigna también una letra a o b minúsculas que enuncian si existe o no ulceración y según el índice mitótico:

- **Grosor del tumor.** Por lo general, los melanomas con un grosor de menos de un milímetro tienen una probabilidad muy pequeña de propagarse. A medida que el melanoma se va haciendo más grueso, presenta mayores probabilidades de propagarse.
- **Índice mitótico:** Un mayor índice mitótico significa tener más células que se están dividiendo lo cual implica que hay más probabilidad de que el cáncer crezca y se propague.
- **Ulceración:** La ulceración es una ruptura en la piel que se encuentra sobre el melanoma. Los melanomas que están ulcerados tienden a presentar un peor pronóstico.

Los posibles valores de la T van de T0 hasta T4b, variando levemente la etapa desde la no existencia de cáncer, hasta que el melanoma mide hasta 4 mm con ulceración

La letra **N** significa que ha habido propagación a los nervios linfáticos cercanos, grupos de células de sistema inmunológico donde el cáncer se propaga primero previamente a la metástasis. Se le asigna un número del 0 al 3 en función de cuanto se han propagado hacia ellos. Pueden contener también letras minúsculas.

Los posibles valores de la N dependen de si se realizó o no una biopsia de ganglio linfático centinela.

La clasificación clínica por etapas de los ganglios linfáticos, la cual se hace sin la biopsia de ganglio centinela, va desde la N0 en la que no ha habido propagación a los ganglios linfáticos a N3 en la que ha habido propagación a 4 o más ganglios linfáticos.

Tras realizar una biopsia de los ganglios linfáticos, se puede determinar la etapa patológica, a la que se le pueden agregar letras minúsculas en algunos casos, informan del tamaño del tumor o de que se han propagado a zonas cercanas al tumor sin llegar a los ganglios.

La letra **M** significa que el melanoma se ha propagado a órganos distantes, es decir, ha habido metástasis. En esta categoría se informa de los órganos alcanzados y de los niveles sanguíneos de una sustancia llamada LDH. La categoría va desde M0 sin metástasis, hasta M2 con metástasis a cualquier órgano distante, M1 es para pulmones y la piel, las minúsculas que las acompañan las diferencian y otorgan información de los niveles LDH.

Una vez definidas las categorías se agrupan para formar las etapas, la asignación de cualquier melanoma a una categoría indicará la etapa en la que se encuentra:

Etapla 0. No existe tumor: **Tis, N0, M0**

Etapla I. Existe tumor con las distintas características:

- IA: **T1a, N0, M0**
- IB: **T1b o T2a, N0, M0**

Etapla II. Existe tumor con mayores riesgos, con las distintas características:

- IIA: **T2b o T3a, N0, M0**
- IIB: **T3b o T4a, N0, M0**
- IIC: **T4b, N0, M0**

Etapla III. Ha habido transmisión a los ganglios linfáticos:

- IIIA: **T1a a T4a, N1a o N2a, M0**
- IIIB: Se aplica en uno de los siguientes casos:
 - **T1b a T4b, N1a o N2a, M0**
 - **T1a a T4a, N1b o N2b, M0**
 - **T1a a T4a, N2c, M0**
- IIIC: Se aplica en uno de los siguientes casos:
 - **T1b a T4b, N1b o N2b, M0**
 - **T1b a T4b, N2c, M0**
 - **Cualquier T, N3, M0**

Etapla IV. Ha habido propagación o metástasis: **Cualquier T, cualquier N, M1(a, b, o c)**

Por lo general, los pacientes con cánceres en etapas menores tienen un mejor pronóstico para una cura o una supervivencia a largo plazo.

2.2.8. Práctica dermatológica

Para poder llevar a cabo un y sistema de prevención adecuado y diagnóstico preciso acerca del cáncer de piel, y más concretamente del melanoma, podemos apoyarnos en la tecnología actual. Existen sistemas que facilitan en gran medida la práctica médica y conceden una mayor seguridad al diagnóstico.

Por parte de los especialistas, como hemos comentado, a la hora de prevenir el desarrollo de células cancerosas en la piel, se realizan registros de los lunares de la piel de los pacientes, para poder realizar un seguimiento de los mismos, posteriormente, o si el paciente demanda un análisis de un lunar en concreto, se procede a realizar un chequeo de los mismos o de los que el especialista considera sospechosos a la hora de desarrollar la enfermedad. Como hemos comentado este procedimiento se denomina dermatoscopia.

2.2.8.1. *Dermatoscopio tradicional*

El análisis de lunares, lesiones y de la superficie en general de la piel del paciente se ha realizado hasta la aparición de sistemas tecnológicos de visualización computarizados por medio de dermatoscopios que contaban únicamente con una serie de lentes y sistemas de iluminación que permitían al especialistas realizar una visualización a tiempo real de la piel del sujeto, sin tener la posibilidad de realizar un registro visual, ni mucho menos un mapeo completo de la misma, así como la capacidad de ver las lesiones ampliadas o con sistemas de fluorescencia que permiten realizar diagnósticos con mayor exactitud. Es por tanto el predecesor de la dermatoscopia digital.



Ilustración 6: dermatoscopio tradicional.

2.2.8.2. Dermatoscopia Digital

La dermatoscopia digital [E8] consiste en un sistema de análisis microscopio asociado a un mapa digital de la piel del paciente que detecta de manera exhaustiva y realiza un seguimiento de lesiones que pueden ser tan mínimas como para ser imperceptibles por el ojo humano. Se basa en tecnologías de captura de imagen microscópica que permite la visualización y ampliación de los lunares o lesiones en una pantalla al mismo tiempo que se toman o almacenan imágenes clínicas para un posterior registro y localización de las mismas. Estas imágenes tomadas son empleadas para realizar seguidamente un seguimiento y estudio de las lesiones y así el especialista podrá detectar cambios en la estructura de la lesión, siguiendo criterios como el ABCD, por ejemplo, y realizar un diagnóstico preciso.



Ilustración 7: dermatoscopia digital.

Concretamente el registro de las lesiones o lunares permite poder observar la evolución de los mismos para poder determinar las posibilidades que tiene de desarrollar un cáncer o si se está desarrollando, siendo este un apartado importante del criterio ABCD.

También es interesante para poder descartar lunares benignos y así no realizar biopsias innecesarias, que serían realizadas como medida de seguridad en caso de no disponer de imágenes de la lesión con una resolución que permita determinar con seguridad que no son malignos, o porque no se haya tenido un registro de su evolución.

Este tipo de sistemas confiere una gran seguridad en pacientes que poseen altos factores de riesgo como son antecedentes de cáncer de piel o gran cantidad de lunares otorgando la posibilidad de un diagnóstico precoz.

La dermatoscopia digital ha supuesto por tanto un importante avance en el diagnóstico precoz del cáncer de piel y el melanoma.

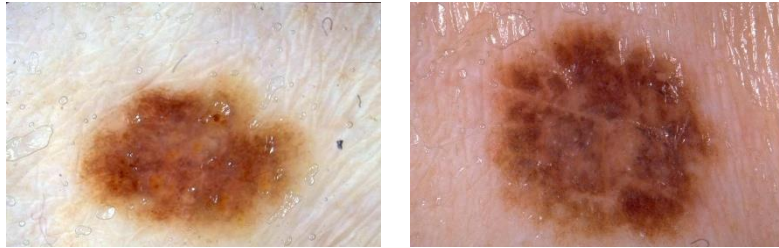


Ilustración 8: Ejemplos de imágenes dermatoscópicas

Por otro lado, por parte de los propios pacientes, si han sido correctamente informados acerca de las prácticas recomendadas, para hacer un seguimiento propio de sus lunares, se realizaran autoexámenes, posean o no factores de riesgo, por prevención.

Existe tecnología para poder apoyar estas prácticas preventivas, tanto las realizadas por especialistas, como las realizadas por el propio paciente. A continuación llevaremos a cabo una muestra con los diferentes sistemas que se emplean actualmente en esta serie de prácticas.

Hay que destacar que existen en muchos casos versiones portátiles de los mismos equipos que se emplean en planta, pero también se desarrollan otros equipos que son destinados a una práctica portable desde el primer momento.

Existen aparatos dermatoscópicos que poseen bases de datos con imágenes histopatológicas tomadas con máquinas del mismo tipo, que se almacenan en nubes de internet y sobre la que los especialistas tienen acceso con el fin de obtener segundas opiniones.

Muchos aparatos además disponen de software de análisis de imágenes utiliza la regla, ABCD o los 7 puntos de Glasgow, para apoyar la clasificación de las lesiones diagnosticadas.

A continuación se describen algunas de las nuevas opciones disponibles como dermatoscopios portátiles

Con la evolución de las tecnologías se está fomentando el uso de aparatos que puedan funcionar como dermatoscopios, que no hayan sido diseñados en origen para tal, pero que sujetos a ciertas modificaciones en hardware y apoyados con un software determinado puedan cumplir con la función dermatoscópica de manera eficaz. Es el caso de cámaras de fotos con lentes específicas para la práctica dermatológica y asociada a un software de visualización [E12].



Ilustración 9: Cámara fotográfica utilizada en dermatoscopia digital

Otro aparato portátil de este estilo que comienza a utilizarse y ha sido desarrollado por parte de empresas del sector, es el Smartphone. Aprovechando la cámara trasera del mismo, a la cual se la acopla lentes y sistemas de iluminación especiales se asocia el uso de la cámara junto con una app con la que gestionar las imágenes tomadas y pedir segundas opiniones a especialistas que dispongan de la aplicación.

Es una alternativa al dermatoscopio digital tradicional que permite una mayor comodidad en la toma de imágenes y un aprovechamiento de la tecnología Smartphone.



Ilustración 10: Teléfono Smartphone adaptado utilizado en dermatoscopia digital

Puede pensarse que es el aparato idóneo para la realización de autoexámenes por parte del paciente en caso de pertenecer a grupos de riesgo pero su precio es elevado y está realmente destinado a especialistas.

Existen también otra serie de aparatos que facilitan el seguimiento de las lesiones son los aquellos que realizan fotografías de los lunares y lesiones del paciente realizando

un mapeo completo del mismo. Así poder realizar un registro y detectar de manera temprana los cambios sufridos en los lunares y prevenir el desarrollo de un posible cáncer.

Pueden tomar fotografías de toda la superficie de la piel del paciente en pocos minutos. Son sistemas que funcionan de manera automática. La cámara fotográfica se mueve de arriba abajo mientras toma fotos con la iluminación correcta [E14].

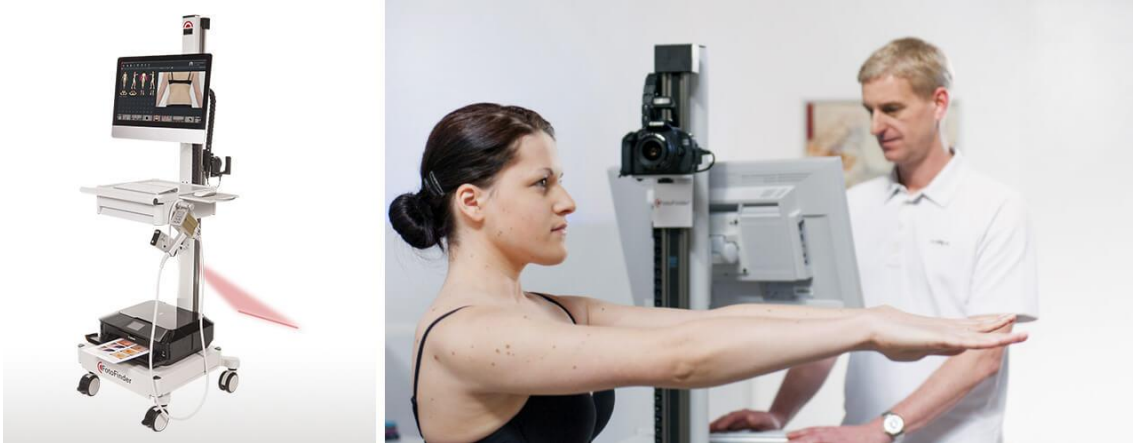


Ilustración 11: Aparato dermatoscópico digital para el seguimiento de las lesiones

2.3. Sistemas de Ayuda al Diagnóstico precoz de Melanomas

Como hemos comentados existen sistemas que sirven de apoyo al diagnóstico precoz de melanomas de manera directa, y se utilizan en el día a día de la práctica médica actual como son los mencionados dermatoscopios digitales, tanto fijos como portátiles, así como los sistemas de seguimiento de lesiones cutáneas; todos ellos son sistemas que apoyan al diagnóstico precoz de lesiones de malignas de la piel y han sido descritas con anterioridad.

A continuación realizaremos una serie de comentarios acerca de algunas de tecnologías, que tienen todas las posibilidades de ser en un futuro cercano elementos primordiales para la práctica médica como sistemas de ayuda al diagnóstico precoz.

Como hemos comentado existen sistemas que aprovechan la tecnología Smartphone para desarrollar aparatos con fines dermatológicos. En apartados anteriores comentamos que dentro de la práctica médica se utilizan teléfonos Smartphone a los que se le adapta un sistema de hardware y software combinado para desarrolla

dermatoscopios portátiles, pero están destinados al uso por parte de especialistas. En cambio en la actualidad se han desarrollado tecnologías que permiten a cualquier usuario, de manera gratuita, hacer un chequeo del estado de sus lesiones capilares. Es el caso de FotoSkin [E17], una app desarrollada en España, que permite el registro de las lesiones por parte del paciente para poder tener un seguimiento de las mismas, y poder mostrarle al especialista la evolución que han seguido sus lesiones. Es una herramienta para apoyar el autoexamen de la piel, y a pesar de que no realiza análisis directo de las lesiones, sirve como herramienta de apoyo al diagnóstico precoz al permitir el seguimiento de las lesiones, y poder tener datos para el correcto desarrollo de criterios como el ABCD, en el que la evolución de la lesión es tan relevante, para realizar un correcto diagnóstico precoz. Además esta app ofrece consejos sobre cuidados, referencias sobre enfermedades y capacidades de asesoramiento para el paciente.

La inclusión de tecnologías de aprendizaje automático para el análisis y reconocimiento de Imágenes es una práctica cada vez más común. En el caso de sistemas de ayuda para el diagnóstico precoz de melanomas, se están desarrollando tecnologías que utilizan también aprendizaje automático. Éste es el caso de la tecnología presentada en la 14 conferencia sobre el Análisis y Reconocimiento de Imágenes presentada en Montreal, Canadá [E16], este pasado Verano, que emplea un software de aprendizaje automático para analizar imágenes de lesiones cutáneas y proporcionar a los médicos datos objetivos sobre biomarcadores reveladores de estos tumores, que son fáciles de tratar cuando el diagnóstico es precoz. Ofrece a los médicos información objetiva sobre las características de la lesión para ayudarles a descartar el melanoma antes de iniciar actuaciones más invasivas como Biopsias. Los desarrolladores enuncian que podría ser una herramienta poderosa para tomar decisiones cara al abordaje del cáncer de piel, y no depender en tanta medida de los exámenes subjetivos de la piel y someter a los pacientes a biopsias para la confirmación de las sospechas. El sistema favorece al diagnóstico precoz basándose en información consistente y cuantitativa sobre cambios en la concentración y distribución de sustancias de la piel que pueden ser claves para el diagnóstico, evitando los retrasos que supone averiguar que les sucede a los pacientes con los medios disponibles hasta la fecha.

También existen tecnologías que emplean aprendizaje profundo (Deep Learning) [E15], para el análisis y reconocimiento de Imágenes de imágenes dermatoscópicas como es el caso de la tecnología desarrollada la compañía tecnológica nipona Kyocera¹ y la Universidad de Tsukuba² desarrolló un sistema para detectar el cáncer de piel en sus fases más tempranas, a partir del análisis de fotografías de pacientes por medio de sistemas de aprendizaje profundo [E18]. El método, aún en fase experimental, ha

¹ <https://www.kyoceradocumentsolutions.es/>

² <http://www.tsukuba.ac.jp/en/>

alcanzado una precisión en sus diagnósticos del 90 por ciento en 4.000 instantáneas analizadas de la piel de los pacientes.

Todo este tipo de tecnologías permiten vislumbrar un futuro esperanzador en la detección temprana del melanoma y en general en la lucha contra el cáncer.

3. Diseño de sistema de ayuda al diagnóstico precoz de melanoma

3.1. Visión General

El sistema de ayuda al diagnóstico precoz de melanomas es un sistema que emplea herramientas de tratamiento digital de la imagen y de aprendizaje automático para que por medio de una base de datos de imágenes dermatoscópicas sea capaz de generar un modelo que clasifique estas imágenes como lesiones benignas o malignas, con la mayor exactitud posible.

El sistema de ayuda al diagnóstico consta de: una base de datos formada por imágenes dermatoscópicas etiquetadas por médicos especialistas y un código programado en el entorno de desarrollo que ofrece el software matemático Matlab.

El planteamiento inicial en el desarrollo del sistema únicamente era emplear las herramientas tradicionalmente utilizadas en la práctica de análisis digital de imágenes y clasificación de las mismas, para cumplir el objetivo fijado de formular un sistema de ayuda al diagnóstico. Por medio del pre-procesamiento de imágenes, una posterior extracción de características diferenciadas manualmente y un aprendizaje maquina supervisado se generarían distintos modelos de clasificación para catalogar las lesiones en benignas o malignas; y por último seleccionar aquel modelo de clasificación que más se ajustara al objetivo inicial.

Una de las técnicas más utilizadas en la actualidad para el análisis digital de imágenes y el tratamiento de datos multimedia es el conocido como aprendizaje profundo, uno de los sistemas más conocidos son las CNN o redes neuronales convolucionales. Éstas obtienen unos resultados excelentes en sistemas de clasificación pero necesitan una cantidad considerablemente elevada de datos para ser entrenadas y una capacidad de cálculo por parte del equipo en el que se desarrolle bastante alta. Ambos son requisitos con los que no se contaban, ya que, la base de datos disponible era reducida y el equipo de desarrollo era un PC portátil de prestaciones normales. Por ello se desestimó la idea de emplear un sistema de aprendizaje profundo completo para realizar nuestra clasificación.

Sin embargo, Matlab cuenta con modelos de redes neuronales convolucionales pre-entrenadas, que permiten trabajar con CNN en equipos de uso doméstico con tiempos de ejecución razonables. Por ello, se planteó la idea de utilizar la red neuronal pre-entrenada para realizar el proceso de extracción de características, adquirirlas en las capas previas a la clasificación del sistema de aprendizaje profundo, y utilizarlas en clasificadores análogos a los utilizados en la extracción de características manual.

De esta manera podríamos contrastar resultados, observar qué sistema de los dos propuestos cumplía mejor con nuestros objetivos, y realizar una comparación de las herramientas tradicionales con las más innovadoras para realizar este tipo de clasificaciones.

Por todo ello, se toma la decisión de utilizar en el desarrollo del sistema de ayuda al diagnóstico un modelo de clasificación que emplea aprendizaje automático tradicional y otro modelo que emplea la combinación del anterior con redes neuronales profundas.

Dentro de éste capítulo realizaremos una descripción de cada uno de los bloques que conforma el sistema de ayuda al diagnóstico para los dos modelos mencionados.

3.2. Descripción del Sistema

3.2.1. Modelo de Aprendizaje Automático

En este apartado describiremos el modelo de aprendizaje máquina desarrollado, haciendo mayor hincapié en cada uno de los descriptores utilizados en el proceso de extracción de características manual.

El modelo consiste en una serie de procesos ilustrados en la Ilustración 12: *(En la ilustración se muestran el bloque de entrenamiento y operación de manera conjunta)*

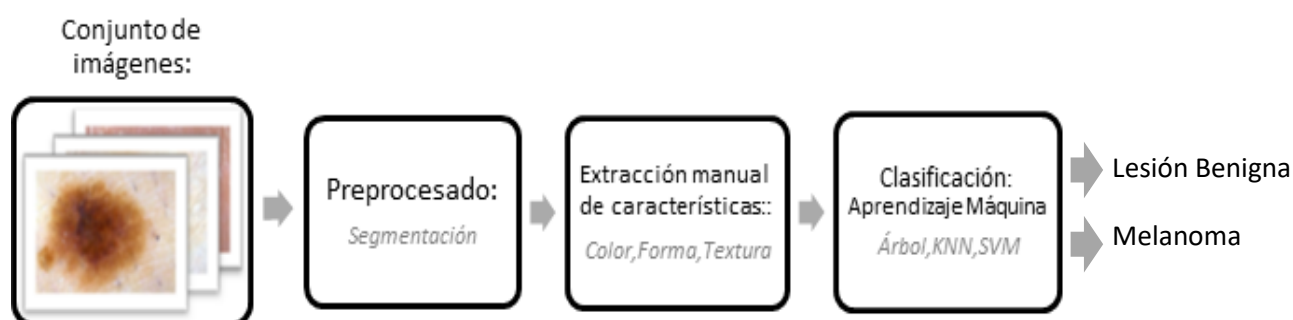


Ilustración 12: Esquema ilustrativo de modelo de aprendizaje automático utilizado

Conjunto de imágenes: Las imágenes dermatoscópicas obtenidas a partir de la base de datos se recopilan en una serie de conjuntos para su posterior introducción en el sistema. En función de la fase de experimentación se emplearán un mayor o menor número de imágenes y se organizarán formando distintos conjuntos de entrenamiento, validación y test para el desarrollo propio del sistema y la realización de pruebas en el mismo.

Pre-procesado: Antes de introducir las imágenes en el proceso de extracción de características diferenciadas manualmente, suele realizarse una preparación de las mismas para desestimar informaciones referentes a la imagen no relevantes para el desarrollo del sistema que incluso pueden mermar sus prestaciones.

En varios procesos similares al implementado suele considerarse que el vello presente en las lesiones es un obstáculo para la correcta clasificación de las lesiones, por lo que suele aplicarse tratamiento digital de la imagen para retirarlo. En otros casos, también se emplean herramientas de filtrado de imagen para eliminar contrastes y altas frecuencias generadas por la porosidad de la piel del individuo. Tras la observación de la base de datos y consideraciones propias se decidió no realizar ninguno de los dos procesos, ya que, en la eliminación del vello y en el filtrado, aparte del efecto deseado, se eliminan aspectos descriptivos de las lesiones que sí pueden llegar a ser de utilidad.

Otro motivo, y quizá el más relevante, es que, la inmensa mayoría de las lesiones disponibles en la base de datos no poseía vello que dificultase la correcta extracción de características, y en caso de presentar vellos, no lo hacían en la zona de la lesión, luego en el proceso de segmentación que describiremos a continuación quedarían eliminados de la misma manera.

Para despejar las dudas acerca de si la eliminación de vello y el filtrado era útil, y para darle al sistema robustez frente a casos en los que se llegara a utilizar muchas lesiones con más vello que las que se disponían se decidió realizar pruebas empleando las dos herramientas de pre-procesado y reafirmando las convicciones iniciales se descubrió que ofrecían peores resultados, por lo que evidentemente se desestimaron.

Una herramienta con la que sí se contó a la hora de preparar las imágenes dermatoscópicas para la extracción de características fue la segmentación. Las imágenes segmentadas permiten centrar la extracción de información sobre la región de interés, desestimando aquellas otras que pueden perjudicar el correcto funcionamiento del sistema.

La base de datos proporcionaba la máscara binaria de las lesiones por lo que fue muy sencillo implementar el proceso de segmentación. Únicamente se debía multiplicar cada una de las lesiones con su respectiva máscara y comprobar que la asociación había sido correcta. El resultado se ilustra en la Ilustración 13:

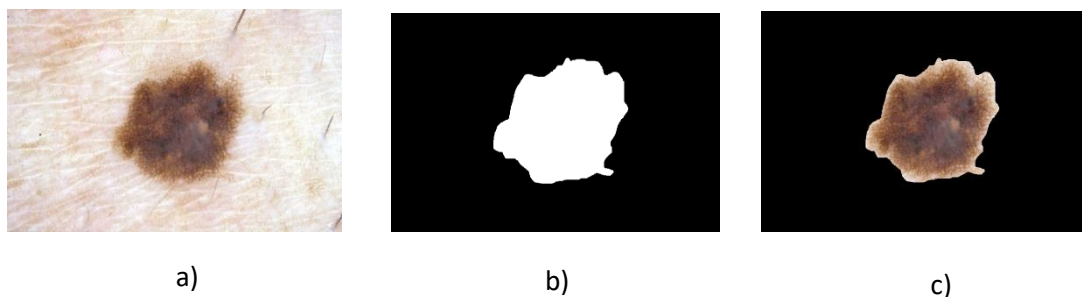


Ilustración 13: a) Imagen Original, b) Máscara, c) Imagen Segmentada.

Extracción de características: precisan de una observación previa de las imágenes a tratar, localizando aquellos aspectos que a juicio del observador (médico) son representativos y descriptivos para poder separar en categorías las distintas imágenes que van a intervenir en el proceso de clasificación.

Los estudios realizados en prácticas dermatológicas son de gran ayuda para determinar esta serie de características visuales propias que permiten disociar en distintas categorías lesiones benignas y malignas, a pesar de ser parecidas a simple vista, por ojos inexpertos.

Si además, añadimos la experiencia de aquellos que realizan prácticas similares a las que pretendemos desarrollar, como es la experiencia del departamento de bioingeniería³ y Cirugía⁴ humana de la universidad de Siena⁵, que logran resultados muy prometedores y son pioneros en el análisis y clasificación de imágenes dermatoscópicas, conseguiremos centrarnos en aspectos clave que caracterizan las imágenes bajo estudio obteniendo una mayor eficiencia y eficacia en el desarrollo.

A la hora de seleccionar qué posibles descriptores son mejores para obtener resultados óptimos en la clasificación de las lesiones, se tomó como referencia los utilizados por experimentos relacionados desarrollados por los departamentos mencionados de la universidad de Siena. Se comenzaron a hacer pruebas con una lista de 35 descriptores que se dividían en tres grandes grupos: Descriptores de color, de forma y de textura.

Por medio de pruebas y resultados obtenidos se fueron incluyendo además de todos estos, algunos relacionados con los del conjunto inicial pero que no pertenecían al mismo en el comienzo, acabando con un conjunto total de 38 descriptores.

Para poder seleccionar de entre todos los descriptores utilizados aquellos que entregaran los mejores resultados y desechar los que añadían información redundante al conjunto, se realizó una serie de pruebas descritas a en este mismo capítulo, pudiendo así localizar los descriptores óptimos.

Más adelante realizamos una descripción de todos los descriptores utilizados en las pruebas, tanto el conjunto inicial como los añadidos posteriormente.

Entrenamiento y clasificación: En el proceso de entrenamiento y posterior clasificación se proponen una serie de modelos de naturaleza diferente, cuyos resultados se pudieran comparar y discutir.

El clasificador de Árbol de decisión resulta intuitivo e interpretable en caso de probar pocos descriptores en fases de desarrollo iniciales.

El clasificador de k-vecinos próximos, emplea un algoritmo que puede ser eficaz en análisis de imágenes y posee parámetros de diseño, como son el número de vecinos o el tipo de distancia empleada, que pueden ajustarse a las necesidades del tipo de imágenes utilizadas.

El clasificador conocido como Máquina de Vectores Soporte (SVM) recibía buenas críticas en distintas documentaciones [E19], y se menciona que el algoritmo que implementa es muy interesante a la par que eficaz. Según los datos recogidos y

³ <https://www.unisi.it/dipartimenti/dipartimento-biotecnologie-mediche>

⁴ <https://www.unisi.it/dipartimenti/dipartimento-scienze-mediche-chirurgiche-e-neuroscienze>

⁵ <https://www.unisi.it/node/3089>

referencias externas se llegó a la conclusión de que sería uno de los que mejores resultados entregaría si no el mejor, por lo que se tomó la decisión de utilizarlo.

En este mismo capítulo realizaremos una descripción más detallada de los distintos clasificadores utilizados tanto para este modelo como para el modelo combinado de aprendizaje automático y profundo.

3.2.1.1. Extracción de Características diseñadas manualmente

A continuación se realiza una breve reseña acerca de los descriptores de imagen empleados en el proceso de extracción de características de cada una de las lesiones.

Estos descriptores se dividirán en tres grupos: aquellos que otorgan información acerca de aspectos referentes al color de la lesión, otros referentes a su tamaño y forma, y los últimos referidos a textura extraídos a partir matriz de coocurrencia de la imagen, que explicaremos más adelante.

Las imágenes introducidas en el proceso de extracción de características son las de la propia lesión multiplicada por su propia máscara binaria, es decir, las lesiones segmentadas, tanto para los descriptores basados en color, como los basados en datos extraídos a partir de la matriz de coocurrencia, basados en textura.

Debido a que no se necesita más que precisamente la forma y tamaño de la lesión, para los descriptores de forma, se emplea directamente la máscara de cada una de las imágenes bajo estudio ya que nos proporciona únicamente la información que se necesita, sin atender a tonalidades o texturas.

3.2.1.1.1. Descriptores basados en color

Como anteriormente comentamos, los descriptores basados en color extraen información relevante sobre algún aspecto que pueda resultarnos útil y ser representativo para poder diferenciar las lesiones malignas de las benignas. Es sabido que existen casos en los que el cáncer de tipo melanoma se manifiesta por medio de cambios en el color de los nevos sobre los que se desarrolla o presenta irregularidades sobre la superficie del mismo (2.2.5) .Trataremos por tanto de extraer información acerca de las tonalidades y proporciones de color de las lesiones para poder contrastar información entre los melanomas y nevos benignos.

3.2.1.1.1.1. *Descriptores basados en media de componentes RGB y HSV*

En un primer momento tomamos la imagen y obtenemos únicamente una componente de la misma, R, G o B, o realizamos un cambio de formato a HSV para posteriormente realizar el mismo proceso. Seguidamente procedemos a extraer el valor medio de la imagen, entendiéndolo como: el valor medio para una componente, en la imagen completa o en el área de la lesión, según las consideraciones que se tengan en cuenta.

El planteamiento inicial era sencillo, emplear una función predefinida por Matlab, para extraer el valor medio de la componente en cuestión. Pero de esta manera se tendría en cuenta los ceros correspondientes a toda la zona anulada por la segmentación o piel no perteneciente a la lesión. Por ello se decide realizar una suma de todos los valores en escala de gris de la imagen y dividir la suma entre el número de píxeles cuyo valor en escala de grises es distinto de cero, es decir, la zona de la lesión que contiene algún valor de la componente en cuestión. Por supuesto, pueden existir zonas de la lesión de valor cero, pero como esto representa la ausencia plena de la componente bajo estudio en esos píxeles se desestiman, ya que además existirían muy pocos píxeles sobre el melanoma que presentarían un cero absoluto.

Representa por tanto el valor medio que toma la componente en los píxeles en que la componente existe, ya que informaciones asociadas al tamaño de la lesión se añadirían con otros descriptores..

Este proceso se realiza igual para cada una de las 6 componentes R, G y B y H, S y V.

3.2.1.1.1.2. *Descriptores basados en varianza de componentes RGB y HSV*

En el caso de la varianza de las distintas componentes presente en las lesiones, tampoco es posible emplear funciones predefinidas en Matlab, ya que tampoco deseamos ponderar aquellas regiones cuyos valores representen la ausencia de la componente bajo estudio tanto dentro como fuera de la lesión.

Se sabe que, la varianza (σ^2) de un conjunto de datos viene expresada como:

$$\sigma^2 = \frac{\sum_{i=0}^{i=n} (x_i - \mu)^2}{n-1}$$

Siendo:

n = El número de datos, en este caso número píxeles considerados.

x_i = Cada valor de la componente en el píxel i -ésimo.

μ = La media de los valores tomados por los píxeles considerados que coincide con el valor extraído en el descriptor anterior.

Teniendo en cuenta todo ello, se implementa un código en Matlab que permite por medio de la expresión anterior, extraer la varianza de las componentes R,G y B y H,S y V de la imagen. Para una descripción más concreta de la implementación en Matlab consultar el Anexo 1.

Debido a que se decide no tener en cuenta los valores no pertenecientes a la lesión, se emplea el sistema anteriormente descrito y por ello no se hace uso del histograma normalizado de la imagen.

Este proceso se realiza igual para cada una de las 6 componentes R, G y B y H, S y V

3.2.1.1.1.3. Descriptores basados en Percentiles componentes RGB

Obtenemos el vector de histograma de las componentes, a continuación sin tener en cuenta los píxeles que representan un 0 absoluto, ya que la inmensa mayoría forma parte de la zona ajena a la lesión, realizamos los porcentajes 10, 20, 40 y 80 que representan el resto de píxeles, y que son pertenecientes a la lesión. Obtenemos la acumulación de píxeles hasta ese porcentaje y devolvemos el nivel de gris que representan, es decir, hasta que nivel de gris representa el 10, 20, 40 y 80 por ciento de los píxeles, y entendiendo esto como los percentiles de la componente.

3.2.1.1.1.4. Descriptores basados en Mediana de componentes RGB

La mediana como parámetro estadístico representa el valor término que divide la distribución de un conjunto de datos ordenados en partes iguales. En nuestro caso representará el valor en escala de gris para cual el histograma de la componente se divida en partes iguales. Éste es un descriptor incluido bajo el criterio del desarrollador.

3.2.1.1.1.5. *Descriptores basados en regiones verdosas de la lesión*

Sabemos que existen ciertas tonalidades que se presentan en las lesiones que han desarrollado melanoma por lo que es interesante extraer información acerca de si estos tonos están presentes en las imágenes bajo estudio y en caso afirmativo el tamaño de las regiones que conforman. En el caso de este descriptor extraeremos la cantidad de píxeles totales que tienen tonalidad verde, entendiendo este matiz como la gama comprendida entre 59/360 y 169/360 sobre la componente H (HSV) representada en la ilustración 14 [E20]:

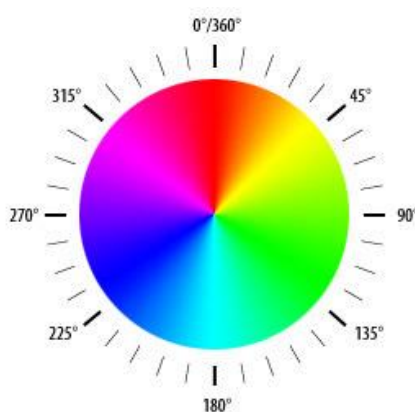


Ilustración 14: Matiz representado en 360 tonalidades sobre H.

Para una descripción detallada de la implementación consultar Anexo 1.

3.2.1.1.1.6. *Descriptores basados en regiones grises-azuladas de la lesión*

Una de las características más descriptivas que se aprecian visualmente en los melanomas es la aparición del denominado velo gris azulado. Suele ser uno de los indicadores dermatológicos más significativos y empleados a la hora de clasificar un melanoma. Se manifiesta como una pigmentación irregular hacia tonos azulados grisáceos formando una película brillante.

Algunas lesiones benignas poseen también tonalidades de este estilo, pero, según datos son un porcentaje muy pequeño con respecto al alto porcentaje de lesiones clasificadas como malignas que poseen esta característica. Por ello los expertos recomiendan extirpar las lesiones que presentan el velo azul-gris.

Debido a todo ello se considera que sería interesante emplear un descriptor que midiera directamente el área o región que posea tonalidades cercanas a la descrita.

El proceso para desarrollar este tipo de descriptor en Matlab es análogo al explicado en el descriptor anterior, únicamente varía el rango de valores de H a considerar, en este caso se tendrán en cuenta valores entre 203/360 y 276/360 que representa según la figura 14.

3.2.1.1.2. Descriptores basados en forma

Esta clase de descriptores aportan información acerca de diferentes aspectos morfológicos que presenta la lesión bajo estudio, por lo que no será necesario tener en cuenta aspectos como el color de la lesión, es decir, será interesante emplear las máscaras de las imágenes disponibles en la base de datos y con las que han sido segmentadas las imágenes para tener una implementación más sencilla de los distintos descriptores sin necesidad de usar imágenes completas en formato HSV y RGB.

Para tres de los cuatro descriptores morfológicos empleados fue posible utilizar una función predefinida en Matlab que proporciona propiedades de medida de regiones de la imagen que introduzcas como parámetro.

3.2.1.1.2.1. Área de la lesión

En la clasificación por etapas del cáncer de tipo melanoma y consecuentemente en los criterios ABCD y de Glasgow [xiv] comentados con anterioridad se hace referencia a tamaños elevados de lesiones generan sospecha en los especialistas así como cambios en el tamaño de los mismos. Será interesante extraer el tamaño de lesión y utilizarla en nuestro sistema.

El procedimiento para obtener el tamaño de la región de interés será sencillo, únicamente se precisará del número de píxeles de la que conforma la lesión a partir de la máscara binaria de la misma. Se obtendrá por medio de funciones predefinidas en Matlab.

3.2.1.1.2.2. Perímetro de la lesión

El concepto y utilidad de este parámetro es similar a la región o área de la lesión. El perímetro de la lesión hace referencia directa al tamaño de la misma, y se representa como el número de píxeles que limitan la región de interés. Se obtendrá de la misma manera, por medio de funciones predefinidas en Matlab.

3.2.1.1.2.3. *Descriptor basado en Circularidad de la lesión*

Para poder obtener la Circularidad de la lesión se desarrolla una función que permite extraer esta propiedad por medio del Área y Perímetro de la misma previamente obtenidos, atendiendo a la siguiente igualdad:

$$Circularidad = \frac{4\pi * \text{Área}}{(\text{Perímetro})^2}$$

3.2.1.1.2.4. *Elipsoidalidad de la lesión*

La elipsoidalidad o excentricidad de la lesión es un parámetro que determina el grado de desviación que posee la lesión con respecto a una circunferencia, es decir, es la contraposición a la circularidad.

Al igual que el área y el perímetro, la Elipsoidalidad o excentricidad de la lesión se obtiene por medio de una función predefinida en Matlab.

3.2.1.1.3. *Descriptores basados en Textura*

Hasta el momento hemos obtenido descriptores que proporcionan información acerca de la morfología y color de las lesiones, pero sabemos que otro aspecto fundamental a la hora de catalogar las mismas como benignas o melanomas, puede ser la textura.

La matriz de coocurrencia permite extraer información acerca de distintos parámetros que describen la textura de las lesiones, sobre todo la referente a estadísticos de segundo orden, que permiten obtener descriptores que los de primer orden no pueden actuando sobre los histogramas de las imágenes.

Este método permite extraer una gran cantidad de información de textura de imagen por la gran variedad de descriptores que es posible obtener de esta matriz, que hacen posible caracterizar con un conjunto de valores cuantificables cada imagen analizada.

A continuación realizaremos una descripción de cómo se conforma la matriz de coocurrencia a partir de cualquier imagen para poder vislumbrar las posibilidades que puede ofrecer a nuestro estudio.

La información que habla acerca de la textura de una imagen está contenida en la relación que existe entre los tonos de gris de la misma. Estas relaciones se especifican,

para pixeles vecinos, en la matriz de coocurrencia espacial para una dirección específica o para todas (0, 45, 90 y 135), para cada par de pixeles.

La matriz de coocurrencia para una dirección determinada se forma de la siguiente manera, en este caso la dirección es hacia la derecha (0°):

Se construye una matriz cuyas filas y columnas depende de las relaciones entre los pixeles vecinos en una dirección determinada para cada uno de los niveles de gris de la imagen, las filas representan el nivel en escala de gris del pixel de referencia, y las columnas representan el nivel del pixel vecino en una dirección preseleccionada, es decir, en cada una de las posiciones de la matriz se enuncia el número de relaciones que se establece en toda la imagen entre un pixel de referencia con un nivel i y el pixel vecino en la dirección preseleccionada j , dando lugar a la diagonal principal en la que el nivel de pixel de referencia y el vecino es el mismo.

Por ejemplo, si una imagen posee 4 niveles de gris conformara una matriz de 4x4 como la siguiente:

Numero de niveles en escala de grises para:		Píxeles vecinos			
		0	1	2	3
Píxeles de referencia	0	(0,0)	(0,1)	(0,2)	(0,3)
	1	(1,0)	(1,1)	(1,2)	(1,3)
	2	(2,0)	(2,1)	(2,2)	(2,3)
	3	(3,0)	(3,1)	(3,2)	(3,3)

Tabla 1: Estructura de matriz de concurrencia.

Como hemos dicho en cada una de las celdas se mostraran el número de relaciones entre pixeles de referencia y vecinos para los niveles de gris especificados entre los mismos según la dirección entre referencia y vecino predefinida. Por tanto, La primera celda debe completarse con la cantidad de veces que ocurre la combinación (0,0). Cuántas veces, en el área de la ventana un píxel con valor de gris igual a 0 (píxel vecino), está situado a la derecha (según la dirección considerada) de otro píxel con valor 0 (píxel de referencia).

Para poder realizar los cálculos posteriores, sin embargo, se necesita que la matriz sea simétrica por lo que se le suma su traspuesta para que quede de esta manera. Otra forma es si se ha realizado la matriz en dirección hacia la derecha sumarle la misma realizada hacia la izquierda y se obtendría el mismo resultado.

Ha de tenerse en cuenta que a la hora de desarrollar las matrices no se tendrán en cuenta aquellas relaciones que no puedan establecerse por pertenecer el pixel de

referencia a un borde. Por ejemplo, no se podrá establecer una relación en dirección 0° (a derechas) entre el pixel de referencia y su pixel vecino si el de referencia no posee más pixeles a la derecha porque delimita el borde de la imagen.

Una vez obtenida la matriz y previamente a los cálculos sobre la misma se realiza una normalización de la matriz de coocurrencia, por medio de la siguiente expresión:

$$C_{i,j} = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}}$$

Siendo:

i = Número de filas

j = Número de columnas

V = Valor de la celda (i, j) en la ventana

$C_{i,j}$ = Probabilidad en la celda i, j

N = Numero de filas o columnas

Como hemos comentado en la diagonal se representan pares de píxeles que no tienen diferencias en su nivel de gris. Esto quiere decir que si estos elementos tienen probabilidades grandes, una vez normalizada la matriz, entonces la imagen no muestra mucho contraste, ya que la mayoría de los píxeles son idénticos a sus vecinos.

Si se suman los valores de la diagonal obtenemos la probabilidad de que un píxel tenga el mismo nivel de gris que su vecino.

Las líneas paralelas a la diagonal separadas una celda, representan los pares de píxeles con una diferencia de 1 nivel de gris. A medida que nos alejamos de la diagonal la diferencia entre niveles de grises es mayor. De la misma manera sumando los elementos separados dos celdas de la diagonal, tenemos los pares de píxeles con dos valores de grises de diferencia, y así de manera sucesiva.

Para poder obtener la matriz de coocurrencia de las imágenes bajo estudio, se emplea una función predefinida en Matlab.

El acceso a los parámetros estadísticos extraídos a partir de la matriz de coocurrencia se realiza por medio de una función definida en contenido descargable de Matlab.

A continuación hacemos una breve descripción de en qué consisten los descriptores estadísticos que informan acerca de la textura de la lesión

3.2.1.1.3.1. *Descriptor basado en Contraste*

El contraste es una medida de las variaciones locales en una imagen. En la matriz de coocurrencia, cuando los valores son altos en la diagonal principal o celdas cercanas a ella, el valor de contraste es bajo, porque quiere decir que existen pocos cambios entre píxeles vecinos. Alcanza un valor alto en caso contrario. Teniendo en cuenta las definiciones anteriormente mencionadas se define:

$$\text{Contraste} = \sum_{i,j} (i - j)^2 \cdot C_{i,j}$$

Como comentamos la obtención del mismo se realiza por medio de una función definida en contenido descargable de Matlab a partir de la matriz de coocurrencia. Para el resto de descriptores basados en textura el procedimiento es similar.

3.2.1.1.3.2. *Descriptor basado en Correlación*

La correlación es una medida de la semejanza que posee la imagen con una réplica de la misma desplazada. Los píxeles más cercanos están más correlacionados entre sí que con píxeles más distantes.

3.2.1.1.3.3. *Descriptor basado en Energía*

La energía es una medida de la homogeneidad local. Cuando la matriz de coocurrencia presenta en sus celdas valores similares la energía presenta valores bajos y al contrario posee valores elevados cuando presenta pocas entradas de gran magnitud. Si todos los píxeles son iguales la energía es mínima. Se define:

$$\text{Energía} = \sum_{i,j} C_{i,j}^2$$

3.2.1.1.3.4. *Descriptor basado en Homogeneidad*

La homogeneidad es la medida opuesta al contraste y se calcula según:

$$\text{Homogeneidad} = \sum_{i,j} \frac{C_{i,j}}{1 + (i - j)^2}$$

3.2.1.1.3.5. *Descriptor basado en Entropía*

Este descriptor esta relacionado con la cantidad de información existente en la imagen. El descriptor en particular alcanza su mínimo cuando todos los elementos de la matriz de coocurrencia son iguales, esto es cuando no hay información. Se calcula mediante la siguiente expresión:

$$Entropía = \sum_{i,j} -C_{i,j} \cdot \log_2[C_{i,j}]$$

3.2.1.1.4. *Selección óptima de características*

La lista de descriptores propuestos para realizar el proceso de extracción de características asciende a 35 contando únicamente con los basados en los utilizados en la universidad de Siena. Si además añadimos otros desarrollados bajo el proceso de experimentación y pruebas la cifra asciende a 38, como son los descriptores de mediana.

Como era de esperar la lista de descriptores es elevada y existe la posibilidad de que varios de ellos entreguen información redundante al sistema. Esto no es interesante pues no ayuda a la obtención de mejores resultados y ralentiza el proceso de extracción y procesamiento de las características. Cuanto mayor sea el número de características a emplear mayor será el número de dimensiones a trabajar en la clasificación y mayor complejidad le dará al sistema.

Para evitar esta redundancia de información en la medida de lo posible y ser efectivos a la hora de encontrar aquellos descriptores que separaran los datos de tal forma que fuera más sencillo clasificar las muestras en melanomas o lesiones benignas se realizaron pruebas experimentales para tratar de encontrar de manera visual cuáles eran los descriptores óptimos. Naturalmente, los que resultaban óptimos para un número de muestras que al principio era bajo, no necesariamente debía serlo cuando eran utilizados en conjuntos de muestras más elevados, por lo que era necesario realizar pruebas para un conjunto de datos superior.

En el capítulo cuatro se explica el proceso llevado a cabo de manera experimental para seleccionar las características que ofrecen los mejores resultados.

Finalmente tras distintas pruebas se llega a la conclusión de que las características que no ayudan a obtener mejores resultados al sistema son: La circularidad, la elipsoidalidad y la región verde de la lesión. Si se realizaban pruebas con todos los descriptores diseñados (38) se obtienen áreas bajo la curva similares a las obtenidas en

la utilización de los 35 descriptores seleccionados, pero los parámetros empleados para dar resultados con carácter clínico eran menos prometedores aunque también parejos por lo que se decide desestimarlos. La intención de retirar ciertos descriptores era mejorar el sistema en la medida de lo posible y reducir el volumen de datos a utilizar pero como comentamos la utilización de todos los descriptores diseñados no obtiene resultados mucho peores.

Por tanto, finalmente se decide que, se emplearán, para el modelo con extracción de características diferenciadas manualmente, 35 descriptores, todos ellos descritos con anterioridad en este mismo capítulo.

3.2.1.1.5. Normalización de las características

El proceso siguiente a la extracción y formalización de la matriz de carteristas para el conjunto de imágenes utilizado, es el proceso de normalización de las mismas.

La matriz de características estaba formada por M filas y N columnas, siendo M el número de características obtenidas y utilizadas, y N el número de imágenes utilizadas en el conjunto.

Como se sabe cada uno de los descriptores entrega magnitudes distintas, algunos por ejemplo obtienen el área en número de píxeles, otros, valores estadísticos, lo cual produce cifras cuyos órdenes de magnitud son distintos. Esto no favorece a que los datos estén estructurados bajo la misma escala y dificulta la interpretación de los mismos en el espacio de características en el proceso de clasificación. Por ello, es interesante normalizar todas las variables para que se encuentren acotadas en un espacio interpretable para el diseñador y que tengan el mismo peso en lo referido a distancias en el espacio de características para un posterior proceso de entrenamiento.

Para el proceso de normalización se empleó una función predefinida en Matlab.

Las variables de las distintas imágenes referidas a la misma característica obtienen media estadística 0 y varianza 1. Este proceso se repite para todas las características, acotándolas a todas en el mismo espacio en las distintas dimensiones.

3.3. Modelo con extracción de características mediante redes neuronales convolucionales

Las redes neuronales convolucionales son modelos de aprendizaje profundo que se entrenan mediante el uso de extensos conjuntos de datos etiquetados y arquitecturas de red neuronal que aprenden directamente a partir de los datos, sin necesidad de una extracción manual de características.

El término “profundo” suele hacer referencia al número de capas ocultas en la red neuronal. Las redes neuronales tradicionales contienen dos o tres capas ocultas, las redes profundas pueden tener hasta 150.

El esquema de una red neuronal convolucional es el que aparece en la Ilustración 15:

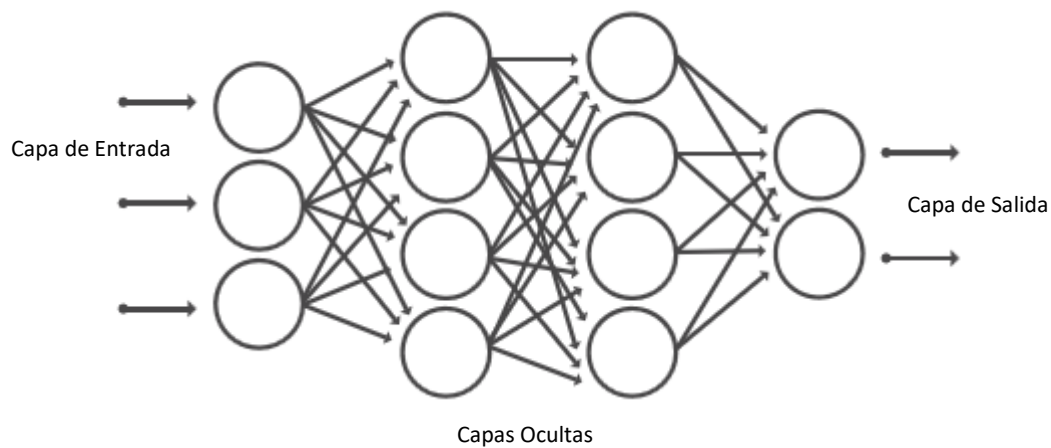


Ilustración 15: Redes neuronales organizadas en capas con una serie de nodos interconectados.

A continuación realizamos una descripción acerca del segundo modelo propuesto para realizar la clasificación de lesiones dermatoscópicas. El modelo consistiría en lo mostrado en la Ilustración 16: *(En la Ilustración 16 se muestra el bloque de entrenamiento y el de operación como uno conjunto)*



Ilustración 16: Diagrama de bloques de modelo en el que se incluye extracción de características por medio de CNN.

Conjunto de imágenes: Para poder desvelar las ventajas que puede tener la inclusión del aprendizaje profundo en nuestro sistema, debemos emplear los mismo conjuntos de imágenes que en modelo inicial, que empleaba en extracción de características diferenciada manualmente.

En esta ocasión las imágenes serán introducidas directamente en el proceso de extracción de características por medio de la red neuronal convolucional, sin pasar por una etapa de pre-procesamiento. La extracción de características en este caso la realiza la red mediante aprendizaje y toma como descriptor la salida de una las últimas capas.

Para decidir si el pre-procesamiento de las imágenes mejoraba el rendimiento del sistema, se realizaron pruebas con tres conjuntos distintos de imágenes: Sin segmentar, segmentadas, y segmentadas sin vello y filtradas. Los mejores resultados fueron obtenidos con las imágenes sin ningún tipo de procesamiento, por lo que se entiende que la red neuronal capta información más útil a partir de la imagen completa. Por tanto, se utilizaron las imágenes sin segmentar, es decir, las originales.

Extracción de características por medio de CNN: [E15] Una red neuronal convolucional extrae directamente las características de las imágenes. Las características relevantes se aprenden mientras la red se entrena con una colección de imágenes, es decir no se diseñan de manera previa. El proceso de aprendizaje para detectar las diferentes características de una imagen lo realizan por medio de decenas o cientos de capas ocultas (depende de la red). Cada capa oculta aumenta el nivel semántico de las características aprendidas. Con el aprendizaje profundo, los pasos de extracción de características y modelización son automáticos, es decir, puede realizar una clasificación deseada si se lo indicas, ya que realiza un aprendizaje completo, se proporcionan datos sin procesar y una tarea que realizar, como clasificar.

El entrenamiento de estas redes requiere un gran volumen de datos y una capacidad de procesamiento que genera velocidades óptimas para equipos no convencionales que emplean GPU; aun así puede tardar días.

Generalmente se usa lo que se denomina “transferencia de aprendizaje” para equipos convencionales con redes existentes pre-entrenadas como AlexNet⁶, que permite además realizar tareas nuevas si introduces datos nuevos, es decir, realizar un nuevo tipo de clasificación. Es el caso utilizado para poder desarrollar el sistema de ayuda al diagnóstico.

Un enfoque que no se utiliza tanto pero que sin embargo resulta útil para nuestro objetivo es utilizar la red para extraer las características como hemos comentado, seleccionando éstas como la salida de una de las últimas capas de la red, pudiendo utilizarlas posteriormente como entrada en un modelo de clasificación como los empleados en el modelo que emplea extracción manual. En nuestro caso se

⁶ <https://es.mathworks.com/help/nnet/ref/alexnet.html>

seleccionara la capa o layer “fc6”, capa completamente conectada que proporciona 4096 características por imagen a la salida.

Clasificación: Para poder contrastar datos y ver qué sistema es más útil para nuestro estudio, y ver si funciona mejor el diseño manual de las características o la extracción por medio de CNN, utilizaremos los mismos clasificadores que en el modelo anterior. Valoraremos por medio de experimentación si cambiar los parámetros de diseño, como pueden ser los vecinos, kernel, o costes introducidos en los clasificadores.

3.4. Clasificación

La clasificación es el proceso por el cual se asigna un dato multimedia a una categoría dada.

En el caso de nuestro sistema empleamos métodos supervisados que requieren de datos etiquetados sobre las categorías establecidas previamente, y nos ofrecerán una clasificación de las muestras de estudio en las categorías que nosotros hemos prestablecido con anterioridad. Por tanto, el objetivo de los sistemas de clasificación es otorgar de una etiqueta o valor sobre una categoría al objeto multimedia, en nuestro caso, imágenes dermatoscópicas.

Para construir los modelos de clasificación empleados para el desarrollo del sistema de ayuda al diagnóstico se utilizan los diferentes procedimientos o fases: fase de entrenamiento, fase de validación y fase de test.

El conjunto de entrenamiento será aquel conjunto de datos con el que, como su nombre indica, será entrenado el algoritmo de clasificación para un posterior examen o test. Se entregarán al algoritmo los distintos datos recogidos por los descriptores de imagen, y tomarán una posición en el espacio de características, este espacio se conforma de n dimensiones, siendo n el número total de características empleadas en la extracción. Del mismo modo, se introducirán en el sistema las etiquetas, o identificadores de cada imagen sobre la pertenencia a una categoría u otra en el proceso de clasificación. De esta manera el algoritmo “aprende” la posición que poseen los datos referentes a un melanoma o lesión benigna tratados en el espacio de características.

Esta información aprendida en el proceso de entrenamiento será la que emplee el modelo para poder clasificar una nueva muestra, ya sea del conjunto de test o de validación, en la clase que considere con mayor probabilidad de pertenecer.

El proceso de validación consiste en emplear una serie de muestras al margen del conjunto de entrenamiento con el fin de realizar un estudio acerca de los parámetros

propios de cada clasificador. De esta manera ver cuáles ofrecen mejores prestaciones para fijarlos y esperar el mejor rendimiento posible para cualquier conjunto de test que se emplee más adelante.

El conjunto de test es el empleado cuando queremos observar las prestaciones de nuestro sistema y es sobre el cual se realiza la clasificación final tras haber entrenado el modelo.

3.4.1. Clasificación por medio de árbol de Decisión

En un árbol de decisión cada nodo del árbol representa un atributo o característica de las muestras, y cada rama representa un posible valor de esa característica. Este algoritmo genera una serie de reglas de decisión, a partir de un conjunto de datos de entrenamiento. Se genera para cada nodo una regla de decisión

Las muestras están representadas por pares tipo característica/valor. Para un funcionamiento óptimo los valores deben ser excluyentes o disjuntos, a pesar de ello se pueden manejar valores numéricos.

Los datos de entrenamiento se introducen en la raíz del árbol, estos se van dividiendo en función de la característica que se seleccione para ramificar el árbol en cada nodo. Las características que se usan para ramificar se eligen en función de una heurística determinada como la proporción de información. La construcción del árbol se detiene cuando alcanza el criterio de parada, como puede ser que no queden más características para dividir los nodos

La función de clasificación empleada tiene valores de salida discretos que pueden ser incluso booleanos [E21].

3.4.2. Clasificación mediante K vecinos más próximos

El algoritmo de k vecinos próximos o k-NN (“nearest neighbourhood”) es un algoritmo que consiste en la comparación directa del nuevo dato entrante (ya sea de test o validación) con los k datos más cercanos al mismo ya registrados con anterioridad en el proceso de entrenamiento del modelo.

El nuevo dato se compara con los k datos más cercanos y se establece la pertenencia del mismo a una categoría u otra por rotación.

La información acerca de la posición que presentan cada uno de los vecinos en el espacio de características y la asignación de esos datos a las categorías etiquetadas se introduce en el proceso de entrenamiento.

La consideración de qué vecinos son más cercanos al nuevo dato de entrada se realiza de acuerdo con la distancia considerada, que puede ser distancia euclídea, coseno etc.

Los datos de entrenamiento son por tanto modelos en sí mismos y será relevante el número de vecinos k a elegir, pues determinará en gran medida el resultado de la clasificación. Generalmente k pequeños generan clasificadores sensibles al ruido y k grandes favorecen la aparición de más clases dentro del conjunto de vecinos.

En el caso del sistema de ayuda al diagnóstico solo contamos con dos categorías, pero si es cierto que la separabilidad de los datos en el espacio de características depende precisamente de la característica en cuestión, por lo que la asignación del valor de k se realizó por medio de experimentación por medio de conjuntos de validación, ya que la dimensión del espacio de características era considerablemente grande para hacerlo visualmente en espacios separados de 2 o 3 dimensiones.

3.4.3. Clasificación mediante máquinas de vectores soporte (SVM)

Se trata de un algoritmo supervisado diseñado en su origen para clasificación binaria E19. Para un conjunto de datos dado, busca la frontera o hiperplano que separe las observaciones de manera óptima. Obtiene los vectores soporte u observaciones de cada categoría cuya distancia al hiperplano sea la mínima de entre todas las observaciones y establece sobre ellas el máximo margen a la hora de trazar el hiperplano.

Debido a que la frontera de separación más sencilla entre muestras es la frontera lineal el algoritmo podrá establecer este tipo de frontera en caso de que las muestras sean linealmente separables para una dimensión determinada. El algoritmo busca la separación lineal de los datos proyectándolos sobre las distintas dimensiones hasta que encuentra aquella en la que los datos son separables. Sin embargo, existen muestras que no son linealmente separables ni se dispone de un espacio de características multidimensional por lo que se recurre a la utilización de núcleos o kernel que permiten realizar proyecciones sobre los datos sobre otras dimensiones y establecer fronteras entre los mismos de grados superiores a las lineales.

Durante la fase de experimentación en el desarrollo del sistema de ayuda al diagnóstico se plantearon distintos modelos basados en SVM ya implementados dentro de la herramienta de software Matlab. Experimentado con los mismos se establecieron distintos kernel que permitieron obtener diferentes resultados en

función de ese mismo núcleo así como entre los distintos modelos planteados como máquina de vectores soporte.

SVM es el algoritmo de clasificación que mejores resultados ha generado en los estudios sobre el desarrollo del sistema. Esto es debido, a que el algoritmo busca directamente realizar una separación entre los datos dentro del espacio de características lo que favorece el establecimiento de fronteras óptimas.

4. Pruebas experimentales y resultados

4.1. Descripción de la Base de datos

La base de datos utilizada en el sistema de ayuda al diagnóstico es una recopilación de una serie de datos médicos referentes a distintas lesiones de distintos pacientes organizados en diferentes carpetas de archivos bajo una serie de patrones. En las distintas carpetas figuran imágenes en formato “jpg”, descripciones y textos en formato “txt”, y una serie de archivos “mat” que contienen diferentes informaciones.

El acceso ordenado a todos los datos médicos de las distintas lesiones y sus imágenes dermatoscópicas correspondientes se realiza por medio de un archivo “mat” cargable en el entorno de desarrollo de Matlab, que contiene en una estructura matricial, las rutas hacia las distintas imágenes, sus máscaras, y los datos médicos referentes a las lesiones organizadas en filas. A continuación lo describiremos más a fondo.

Una vez cargado este archivo en el entorno de desarrollo, se generaba una estructura que contenía la matriz brevemente descrita, que poseía 724 lesiones y 15 datos médicos asociados a cada lesión, es decir, por cada fila aparecía una lesión distinta, y en cada una de las columnas un tipo de dato médico.

En la matriz mencionada anteriormente cada lesión posee los siguientes campos:

Columna 1: Número o nombre del caso o lesión.

Columna 2: Ruta de acceso hacía la imagen de la lesión en la carpeta de archivos de la base de datos.

Columna 3: Ruta de acceso hacía la máscara binaria de la lesión en la carpeta de archivos de la base de datos.

Columna 4: tipo de lesión dermatológica diagnosticada entre distintos tipos de lesión benigna y melanomas: Clark nevus, Spitz Nevus, Reed/Spitz Nevus, Blue Nevus, Melanoma, Congenital nevus, entre otros.

Columna 5: Patrón global de la lesión. El patrón puede ser homogéneo, reticular o multicomponente entre otros.

Columna 6: Pigmentación de la lesión. Se muestra si la lesión presenta una pigmentación regular difusa o localizada si no es así se marca como “ausencia”.

Columna 7: Hipo-pigmentación de la piel del paciente. Se muestra si la lesión presenta una Hipo-pigmentación focal o multifocal, si no es así se marca como “ausencia”.

Columna 8: Red de pigmentación. Se muestra si la lesión presenta una red de pigmentación típica o atípica, si no es así se marca como “ausencia”.

Columna 9: Velo azul-gris. Se muestra si presenta o si hay ausencia de ello.

Columna 10: Estructura vascular. Regresión interna, punteada, curvada, etc.

Columna 11: Glóbulos de la lesión. Regulares, irregulares o ausencia.

Columna 12: Manchas: Regulares, irregulares o ausencia

Columna 13: Estructura de regresión. Áreas azules, áreas blancas, combinaciones, etc.

Columna 14: Color de la lesión. Los distintos colores que presenta la lesión.

Columna 15: Localización. Zona corporal en la que se localiza la lesión.

La Ilustración 17 proporciona un ejemplo intuitivo de la matriz mencionada:

724x1 struct with 15 fields

...	CaseNumber	ImPath	MaskPath	Diagnosis	GlobalPattern	Pigmentation	Hypopigmentation
64	'4A34'	'../derm_db_...	'../derm_db_...	'Clark nevus'	'reticular pattern'	'localized irregul...	'multifocal'
65	'4A345346'	'../derm_db_...	'../derm_db_...	'Clark nevus'	'multicomponen...	'absent'	'multifocal'
66	'4A347348'	'../derm_db_...	'../derm_db_...	'Clark nevus'	'globular pattern'	'absent'	'absent'
67	'4A349350'	'../derm_db_...	'../derm_db_...	'Clark nevus'	'reticular pattern'	'absent'	'absent'
68	'4A355356'	'../derm_db_...	'../derm_db_...	'Clark nevus'	'reticular pattern'	'diffuse regular'	'absent'
69	'4A357358'	'../derm_db_...	'../derm_db_...	'Reed/Spitz ...'	'multicomponen...	'diffuse irregular'	'absent'
70	'4A361362'	'../derm_db_...	'../derm_db_...	'Clark nevus'	'starburst pattern'	'diffuse regular'	'absent'
71	'4A365366'	'../derm_db_...	'../derm_db_...	'Clark nevus'	'reticular pattern'	'absent'	'absent'
72	'4A370371'	'../derm_db_...	'../derm_db_...	'Clark nevus'	'reticular pattern'	'absent'	'absent'
73	'4A376377'	'../derm_db_...	'../derm_db_...	'Congenital ...'	'multicomponen...	'absent'	'multifocal'
74	'4A380381'	'../derm_db_...	'../derm_db_...	'Clark nevus'	'reticular pattern'	'absent'	'focal'
75	'4A386387'	'../derm_db_...	'../derm_db_...	'Clark nevus'	'reticular pattern'	'absent'	'absent'
76	'4A388389'	'../derm_db_...	'../derm_db_...	'Melanoma'	'unspecific patte...	'diffuse irregular'	'absent'
77	'4A390391'	'../derm_db_...	'../derm_db_...	'Clark nevus'	'starburst pattern'	'diffuse irregular'	'absent'
78	'4A394395'	'../derm_db_...	'../derm_db_...	'Reed/Spitz ...'	'starburst pattern'	'diffuse regular'	'absent'
79	'4A398399'	'../derm_db_...	'../derm_db_...	'Reed/Spitz ...'	'starburst pattern'	'diffuse irregular'	'absent'
80	'4A400401'	'../derm_db_...	'../derm_db_...	'Reed/Spitz ...'	'starburst pattern'	'diffuse irregular'	'absent'
81	'4A404405'	'../derm_db_...	'../derm_db_...	'Reed/Spitz ...'	'starburst pattern'	'diffuse regular'	'absent'
82	'4A406407'	'../derm_db_...	'../derm_db_...	'Reed/Spitz ...'	'starburst pattern'	'diffuse regular'	'absent'
83	'4A412413'	'../derm_db_...	'../derm_db_...	'Clark nevus'	'reticular pattern'	'absent'	'multifocal'

Ilustración 17: Ejemplo de sección de matriz de datos médicos de las lesiones

Para el desarrollo del sistema de ayuda al diagnóstico y la realización de pruebas con el mismo, he obtenido las imágenes dermatoscópicas a partir de la matriz mostrada en la ilustración 17.

Por medio de un bucle se almacenan todas las imágenes deseadas en una matriz compuesta por imágenes, accediendo al parámetro en el que aparecen las rutas de las imágenes por medio de una función de lectura.

Es en este punto en el que aparece el primero de los problemas, ya que, a pesar de que todas las imágenes poseían las mismas dimensiones, existían imágenes dañadas que carecían de una fila o columna de píxeles, debido a lo que parecen errores en la base de datos. Esto hacía que se debiera desestimar este tipo de imágenes a la hora de trabajar y dificultaba la importación de las mismas, sin mencionar el tiempo perdido en localizar el problema.

Una gran ventaja, a la hora de realizar el pre-procesamiento de las imágenes, fue que la base de datos disponía de las máscaras binarias de las lesiones, lo que supuso un ahorro de tiempo en el proceso de segmentación de las imágenes al no tener que obtenerlas. En un caso real, será necesario diseñar un sistema de segmentación automática previo.

El proceso de importación de las máscaras al entorno de desarrollo es similar al de las imágenes. En este caso se accede a la columna que contiene la ruta de acceso hacia las máscaras en la carpeta de archivos. Empleando una función se carga la máscara binaria, almacenada en formato “mat”, formato exclusivo de Matlab y que permite guardar datos referentes al entorno de desarrollo.

Este proceso también tuvo una serie de dificultades ya que dentro de la estructura referente a la matriz de datos médicos aparecían errores de notación dentro de la ruta que daba acceso a la máscara. El propio archivo de la máscara, estaba enunciado en formato “jpg”, archivo incompatible con la función de carga mencionada, lo que produjo que hubiera que ir cambiando a mano esta terminación por “mat”.

La base de datos ha sido adquirida por el departamento de Teoría la Señal de la Universidad. El departamento ha permitido que sea utilizada una copia de manera exclusiva para hacer este Trabajo de Fin de Grado, en el momento de conclusión y entrega del mismo será eliminada.

4.1. Ensayos preliminares con descriptores diseñados manualmente

Para el sistema de ayuda al diagnóstico, en las primeras pruebas, se emplearon un conjunto de diez imágenes balanceado, con cinco lesiones benignas y cinco melanomas. Este mismo conjunto se utilizó tanto para entrenamiento como para test, y así poder comprobar que el entrenamiento llevado a cabo por los distintos clasificadores empleados era correcto, observando áreas bajo la curva iguales a 1, y realizando pruebas con todos los descriptores diseñados, pudiendo de esta manera, comenzar a tener nociones de cuáles de las características extraídas de las imágenes separaban en mayor medida los datos. Estas pruebas no serían indicativas de la precisión del clasificador pero servirían para empezar con buen pie, e ir avanzando poco a poco en el desarrollo del sistema de clasificación.

En pruebas posteriores, se aumentó el volumen de los conjuntos de datos: 76 imágenes para entrenamiento, 13 para validación y 13 para test, todos ellos balanceados. Por medio del conjunto de validación se fijaron los parámetros de diseño de los clasificadores hasta que se lograron las mejores prestaciones del sistema. Posteriormente se realizaron pruebas con el conjunto de test inicial de 13 imágenes.

Observando la separabilidad de los datos a simple vista, por medio de nubes de puntos y las propias tablas de características se lograron prestaciones prometedoras para este conjunto de datos, con áreas bajo la curva de 0,97. Los resultados obtenidos eran a pesar de ser buenos, únicamente indicativos de que el proceso se desarrollaba correctamente, ya que un volumen de datos de 102 muestras era muy reducido. En estos casos los modelos utilizados se ajustan en gran medida al conjunto de entrenamiento y no generalizan correctamente para cualquier conjunto, ni garantizan robustez frente a posteriores testeos.

En las últimas pruebas llevadas a cabo, el aumento en el volumen de datos, con 314 imágenes de entrenamiento y 104 de test, ambos balanceados, con los descriptores seleccionados como óptimos a simple vista en fases anteriores, el sistema pierde prestaciones, por lo que se procede a realizar un análisis de cuáles son los descriptores óptimos, para determinar aquellas características que ofrecen mayor separabilidad para este mayor volumen de datos, y que por tanto, ofrecerían mejores prestaciones. Este proceso además debería reducir el volumen de datos total a introducir en el proceso de clasificación, en cuanto a características de las imágenes se refiere.

Se realiza, por tanto, para este volumen de imágenes, diferentes procesos de entrenamiento y test para cada uno de los descriptores y se hace un registro de los resultados. Posteriormente, una vez seleccionados aquellos que ofrecen mejores prestaciones, se combinan y se realiza otra serie de procesos entrenamientos y test

para estas combinaciones. De nuevo se realiza un registro de los resultados como anteriormente y en función de los mismos se continúa realizando pruebas y combinaciones hasta que los resultados no superen el mejor de los obtenidos. Este es un procedimiento tedioso, pero eficaz, ya que permite obtener resultados óptimos, desestimando aquellos descriptores que ofrecen datos que no mejoran las prestaciones del sistema.

Tras el proceso, se logra una mejora en las prestaciones iniciales, sin embargo, no se obtienen resultados tan prometedores con segundo conjunto de datos, que era más reducido. Se decide finalmente, que éste sería el último conjunto de datos que se emplearía a la hora de realizar pruebas, ya que la base de datos disponible contaba con un total de 724 lesiones de las cuales 210 eran malignas y estaban ya en su totalidad en los conjuntos empleados, además, resultaba más cómodo emplear conjuntos de datos balanceados y por ello se decide no aumentar el volumen de datos de trabajo pues únicamente se habría introducido un mayor número de lesiones benignas en el mismo.

4.2.Resultados

Los resultados se presentarán en los siguientes términos:

$$\text{Precisión} = \frac{Tp}{Tp+Fp}$$

$$\text{Sensibilidad/Recall} = \frac{Tp}{Tp+Fn}$$

$$\text{Especificidad} = \frac{Tn}{Tn+Fp}$$

$$F1_Score = \frac{2 \cdot \text{Precisión} \cdot \text{Recall}}{\text{Precisión} + \text{Recall}}$$

Ciertos positivos (TP): Melanomas reales clasificados correctamente como melanomas

Falsos positivos (FP): Lesiones benignas clasificadas incorrectamente como melanomas

Ciertos negativos (TN): Lesiones benignas clasificadas correctamente como benignas

Falsos negativos (FN): Lesiones malignas clasificadas incorrectamente como benignas

Debido al carácter y objetivo clínico que posee el sistema desarrollado utilizamos los términos *Especificidad* y *Sensibilidad*, que son empleados en la práctica médica.

En el caso de detección de lesiones benignas y melanomas estos parámetros se refieren a lo siguiente: La especificidad se refiere a la cantidad de lesiones benignas que han sido correctamente clasificadas como benignas; mientras que la sensibilidad se refiere a la cantidad de melanomas que han sido correctamente clasificados como melanomas de entre todos los casos reales de melanoma bajo estudio, es decir, tanto los que se ha acertado en su clasificación, como los melanomas que han sido clasificados como benignos.

La precisión se refiere a la cantidad de melanomas que se han clasificado correctamente entre todas las lesiones clasificadas como melanomas es decir, aquellas que son benignas pero se ha fallado en su clasificación, y aquellas que son malignas y se ha clasificado correctamente.

El *F1-score* o valor F1 es un término estadístico que representa la media armónica que combina los valores de precisión y de sensibilidad anteriormente mencionados.

A continuación mostraremos las anteriores magnitudes junto a otra serie de resultados para los dos modelos empleados con los distintos clasificadores.

4.2.1. Descriptores diseñados manualmente

Posteriormente al proceso de selección de características por medio de los métodos mencionados en capítulos anteriores se llega a la conclusión de que el número de descriptores que ofrecen los mejores resultados asciende a 35 de un total de 38.

Los resultados para los 35 descriptores diseñados manualmente son los que se mostrarán a continuación para los tres modelos de clasificación seleccionados.

La política de costes que se realizó en este proceso pretendía que los modelos de clasificación obtuvieran una sensibilidad aceptable en cuanto a términos médicos se refiere por encima de la especificidad lograda. En el desarrollo de los resultados se detalla la política de costes para cada uno de los modelos de clasificación.

4.2.1.1. Resultados de clasificación por medio de Árbol de Decisión

El único parámetro de diseño seleccionado para el clasificador por medio de Árbol de Decisión es el coste. Se le da un coste superior a la no detección de un melanoma frente a la no detección de una lesión benigna con una relación 2:1;

El área bajo la curva es la siguiente:

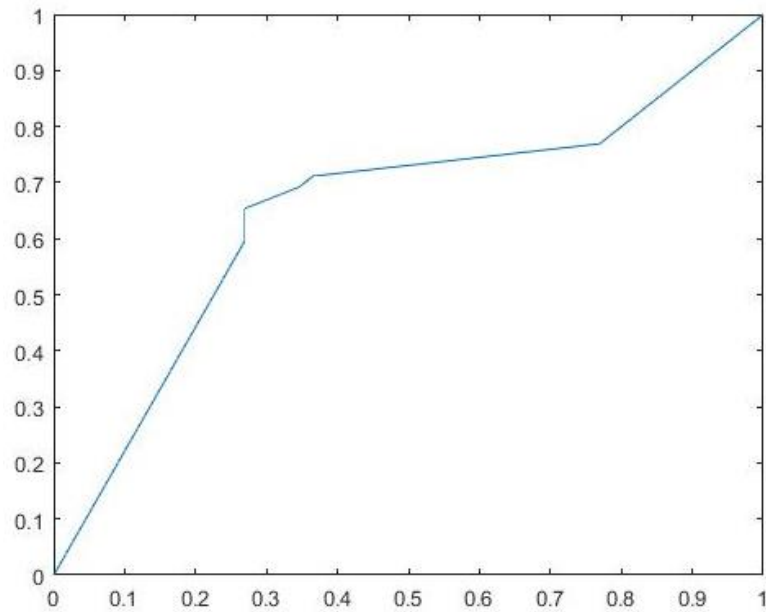


Ilustración 18: Área bajo la curva de clasificador por medio de Árbol de Decisión

A continuación en la Tabla 2 mostramos los resultados obtenidos en el test:

Área bajo la curva	0.67
Precisión	0.66
Sensibilidad	0.69
Especificidad	0.65
Valor F1	0.68

Tabla 2: Resultados para el clasificador por medio de Árbol de Decisión

A continuación en la Tabla 3 mostramos la matriz de confusión con la que se han generado los resultados anteriores excepto el área bajo la curva:

Matriz de confusión Árbol de Decisión		Clasificación	
		Melanoma / P	Benigno/ N
Clase Real	Melanoma/ P	Tp:36	Fn:16
	Benigno/ N	Fp:18	Tn:34

Tabla 3: Matriz de confusión para el clasificador por medio de Árbol de Decisión

4.2.1.2. Resultados de clasificación por K vecinos próximos (K-NN)

Los distintos parámetros de diseño seleccionados para el clasificador por K-NN son: el coste, el número K de vecinos y el tipo de distancia empleada en el espacio de características. Se le da un coste superior a la no detección de un melanoma frente a la no detección de una lesión benigna con una relación 2:1, como en el caso anterior. El número de vecinos óptimo obtenido con conjunto de validación es K=3. La distancia que ofrece mejores resultados según experimentos en el proceso de validación es la distancia coseno.

El área bajo la curva es la siguiente:

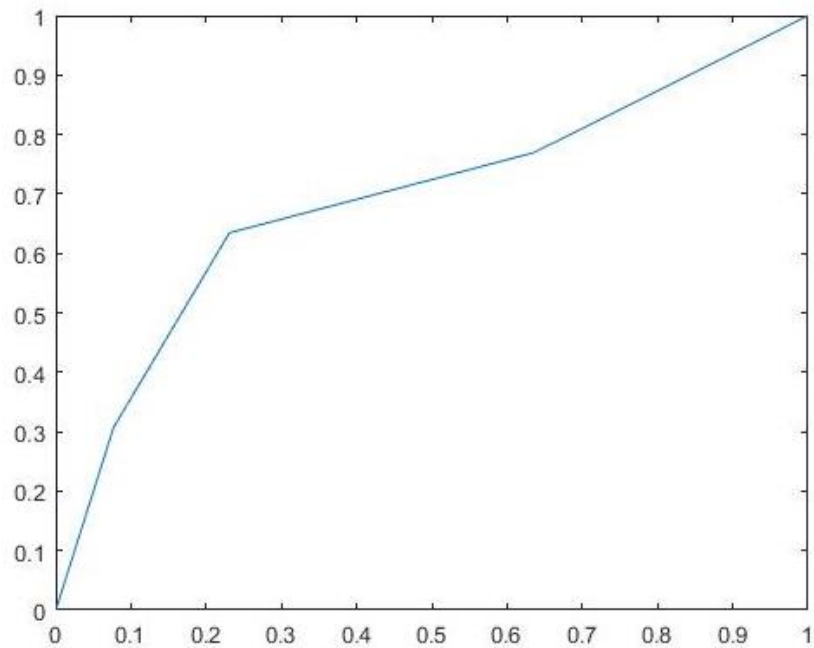


Ilustración 19: Área bajo la curva de clasificador por K-NN

A continuación en la Tabla 4 mostramos los resultados obtenidos en el test:

Área bajo la curva	0.69
Precisión	0.73
Sensibilidad	0.63
Especificidad	0.77
Valor F1	0.68

Tabla 4: Resultados para el clasificador por medio K-NN

A continuación en la Tabla 5 mostramos la matriz de confusión con la que se han generado los resultados anteriores excepto el área bajo la curva:

Matriz de confusión K-NN		Clasificación	
		Melanoma / P	Benigno/ N
Clase Real	Melanoma/ P	Tp:33	Fn:19
	Benigno/ N	Fp:12	Tn:40

Tabla 5: Matriz de confusión para el clasificador por K-NN

4.2.1.3. Resultados de clasificación por Máquinas de Vectores Soporte (SVM)

Los distintos parámetros de diseño seleccionados para el clasificador por medio de Máquinas de Vectores Soporte son: el coste y el tipo de kernel o núcleo utilizado. Se le da un coste superior a la no detección de un melanoma frente a la no detección de una lesión benigna con una relación 2:1, como en los dos casos anteriores. El tipo de núcleo utilizado es lineal, se ha fijado por procesos de validación.

El área bajo la curva es la siguiente:

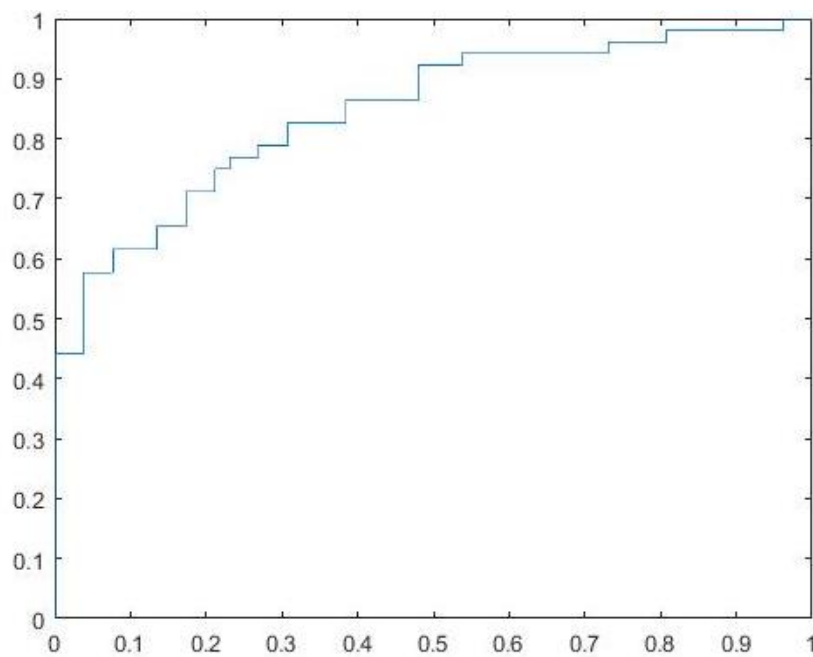


Ilustración 20: Área bajo la curva de clasificador por SVM

A continuación en la Tabla 6 mostramos los resultados obtenidos en el test:

Área bajo la curva	0.85
Precisión	0.71
Sensibilidad	0.83
Especificidad	0.65
Valor F1	0.76

Tabla 6: Resultados para el clasificador por medio SVM

A continuación en la Tabla 7 mostramos la matriz de confusión con la que se han generado los resultados anteriores excepto el área bajo la curva:

Matriz de confusión SVM		Clasificación	
		Melanoma / P	Benigno/ N
Clase Real	Melanoma/ P	Tp:43	Fn:9
	Benigno/ N	Fp:18	Tn:34

Tabla 7: Matriz de confusión para el clasificador por medio SVM

4.2.2. Descriptores obtenidos por medio de CNNs

Se decide que sería interesante, a la hora de comparar el modelo con extracción de características por medio de CNN con el anterior, que utilizaran ambos la misma política de costes. Por ello para estos clasificadores también se emplea un modelo de entrenamiento con coste superior en la no detección de un melanoma una relación 2:1 frente a la no detección de la benigna.

4.2.2.1. Resultados para clasificación por medio de Árboles de Decisión

El único parámetro de diseño es el coste y es el definido con anterioridad. El área bajo la curva se representa en la siguiente figura:

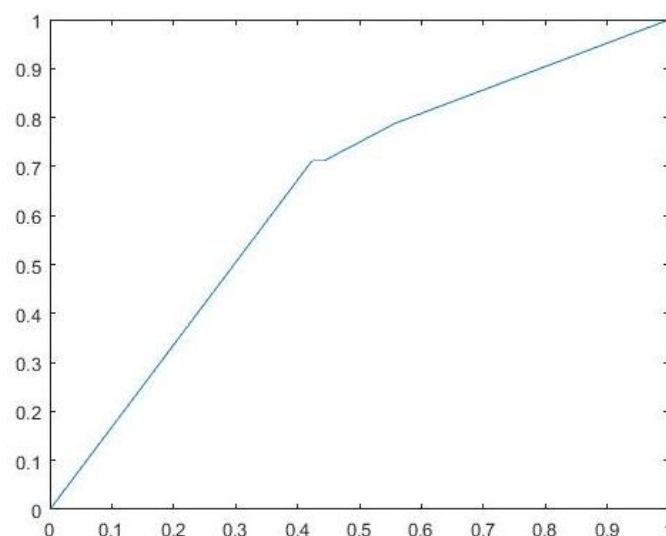


Ilustración 21: Área bajo la curva del clasificador basado en Árbol de Decisión para la extracción de características por CNN

A continuación en la Tabla 8 mostramos los resultados obtenidos en el test:

Área bajo la curva	0.65
Precisión	0.62
Sensibilidad	0.71
Especificidad	0.56
Valor F1	0.66

Tabla 8: Resultados para el clasificador por medio de Árbol de Decisión para la extracción de características por CNN

A continuación en la Tabla 9 mostramos la matriz de confusión con la que se han generado los resultados anteriores excepto el área bajo la curva:

Matriz de confusión Árbol de Decisión		Clasificación	
		Melanoma / P	Benigno/ N
Clase Real	Melanoma/ P	Tp:37	Fn:15
	Benigno/ N	Fp:23	Tn:29

Tabla 9: Matriz de confusión para el clasificador por medio de Árbol de Decisión para la extracción de características por CNN

4.2.2.2. Resultados para clasificación por medio de K-NN

Los parámetros utilizados son los mismos que los utilizados en el clasificador basado en K-NN del modelo de extracción de características diferenciadas manualmente.

El área bajo la curva es:

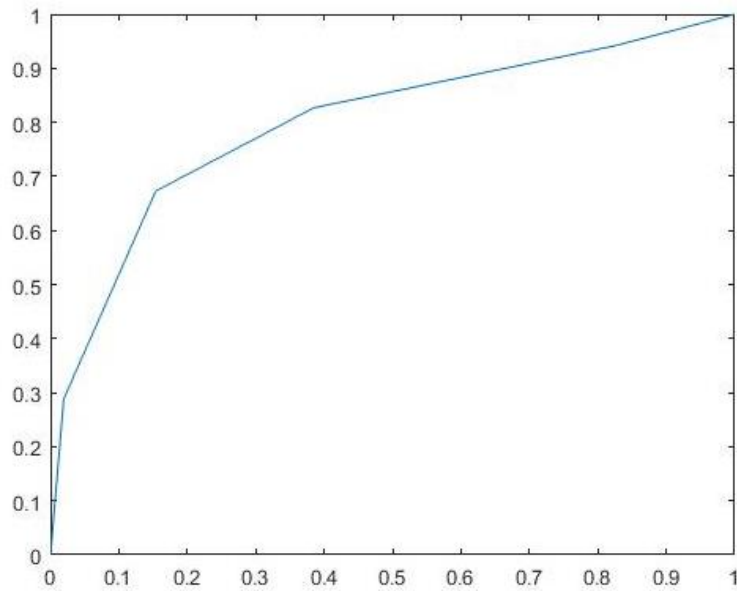


Ilustración 22: Área bajo la curva de clasificador por K-NN para la extracción de características por CNN

A continuación en la Tabla 10 mostramos los resultados obtenidos en el test:

Área bajo la curva	0.79
Precisión	0.68
Sensibilidad	0.83
Especificidad	0.61
Valor F1	0.75

Tabla 10: Resultados para el clasificador por medio K-NN para la extracción de características por CNN

A continuación en la Tabla 11 mostramos la matriz de confusión con la que se han generado los resultados anteriores excepto el área bajo la curva:

Matriz de confusión K-NN		Clasificación	
		Melanoma / P	Benigno/ N
Clase Real	Melanoma/ P	Tp:43	Fn:9
	Benigno/ N	Fp:20	Tn:32

Tabla 11: Matriz de confusión para el clasificador por K-NN para la extracción de características por CNN

4.2.2.3.Resultados para clasificación por medio de Máquina de Vectores Soporte (SVM)

Los distintos parámetros de diseño seleccionados para el clasificador por medio de Máquinas de Vectores Soporte son igual que en el modelo de extracción de características diferenciadas manualmente: el coste y el tipo de kernel o núcleo utilizado.

El tipo de núcleo utilizado es lineal, se ha fijado por procesos de validación, como en casos anteriores. El coste es el comentado.

El área bajo la curva es la siguiente:

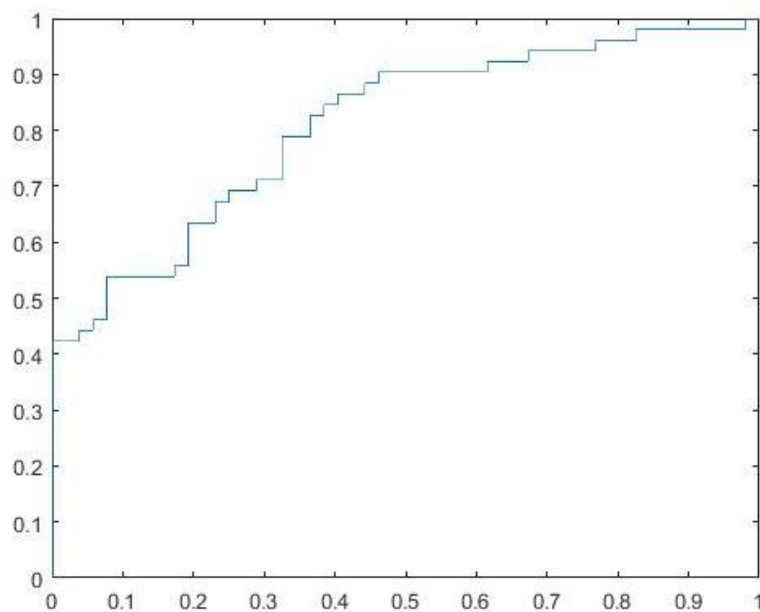


Ilustración 23: Área bajo la curva de clasificador por SVM en modelo de extracción de características por medio de CNN

A continuación en la Tabla 12 mostramos los resultados obtenidos en el test:

Área bajo la curva	0.81
Precisión	0.68
Sensibilidad	0.79
Especificidad	0.63
Valor F1	0.73

Tabla 12: Resultados para el clasificador por medio SVM en modelo de extracción de características por medio de CNN

A continuación en la Tabla 13 mostramos la matriz de confusión:

Matriz de confusión SVM		Clasificación	
		Melanoma / P	Benigno/ N
Clase Real	Melanoma/ P	Tp:41	Fn:11
	Benigno/ N	Fp:19	Tn:33

Tabla 13: Matriz de confusión para el clasificador por medio SVM en modelo de extracción de características por medio de CNN

4.2.3. Errores cometidos

A continuación mostramos tres ejemplos de lesiones benignas que han sido clasificadas como malignas tanto en el sistema de extracción de características diseñadas manualmente , como en el de extracción por medio de CNNs, en el modelo de clasificador basado en SVM que es el que mejores resultados ofrece en ambos sistemas:



Ilustración 24: Ejemplos de lesiones benignas clasificadas como melanomas.

Las tres lesiones son benignas, pero cada una de ellas presenta características morfológicas y visuales distintas. La imagen de la izquierda presenta contornos

irregulares y poca homogeneidad en las tonalidades y texturas en toda la región de la lesión, características propias de los melanomas.

La lesión central y sobre todo la de la derecha presentan tonalidades azuladas muy características de los melanomas. La lesión de la derecha presenta además poca homogeneidad en las texturas a pesar de que no es tan acusada como en el caso de la izquierda. La imagen central es bastante homogénea pero con tinciones azules como comentábamos. Las tres presentan tamaños considerables en comparación con otras lesiones que es otra característica que puede inducir a error. Todas las características mencionadas favorecen en estos casos la aparición de falsas alarmas para los dos sistemas de extracción de características diseñados.

En las tres imágenes que se muestran en la siguiente figura aparece el caso contrario al descrito con anterioridad, son lesiones malignas que han sido clasificadas como benignas en ambos sistemas empleados para la extracción de características para el modelo de clasificación que mejores resultados otorga como es el basado en SVM:



Ilustración 25: Ejemplos de melanomas clasificados como lesiones benignas.

En cada una de las tres lesiones no aparecen tinciones cercanas a tonos azulados. El tamaño del melanoma central es bastante reducido. El melanoma de la derecha presenta homogeneidad en las texturas que lo componen. El melanoma de la izquierda presenta mayor contraste en la superficie y es de un tamaño elevado pero posee tinciones más propias de lesiones benignas. Las características mencionadas son propias de lesiones benignas por lo que podemos hacernos una idea de por qué han fallado ambos diseños.

Puede acusarse uno de los fallos a la no utilización de herramientas morfológicas como los descriptores de circularidad y elipsoidalidad ya que algunas lesiones mostradas presentan con respecto a estas características aspectos propios de los grupos en los que se les han clasificado como erróneos. Como se comentó fueron rechazadas por entregar aunque resultados muy parejos, en lo referente a área bajo la curva a los obtenidos finalmente, peores resultados en los parámetros utilizados para práctica médica por lo que se desestima su utilización. El extractor de características por medio de redes neuronales convolucionales sí tiene la posibilidad de extraer información acerca de la excentricidad de las lesiones y ha fallado en lesiones como las de este

estilo por lo que no podemos adjudicar el fallo a la no utilización de los descriptores mencionados y sí probablemente al resto de características descritas.

Las imágenes mostradas son solo algunos ejemplos prácticos para poder mostrar las irregularidades enunciadas, ya que concretamente estas se clasificaban mal para ambos modelos de extracción de características, para poder extraer las siguientes conclusiones se visualizó además un mayor número de imágenes erróneas para cada uno de los casos.

En general Los errores cometidos en la clasificación de lesiones benignas como melanomas para ambos casos se localizan en lesiones que poseen tonalidades azules, presentan texturas poco homogéneas e irregularidades morfológicas y tamaños elevados.

En el caso contrario los errores cometidos en la clasificación de lesiones malignas como lesiones benignas se producen en aquellas que poseen tonalidades propias de lunares sanos, sin tinciones azules, que presentan texturas homogéneas y con pocas irregularidades y que no presentan tamaños elevados.

5. Conclusiones y Líneas Futuras

5.1. Conclusiones Obtenidas

Finalmente, tras la visualización de los resultados obtenidos para ambos modelos concluimos lo siguiente.

Para el modelo con extracción de características diseñadas manualmente, observando los resultados de los distintos clasificadores empleados podemos comprobar que modelo de clasificación que mejores prestaciones ofrece es el de Máquina de Vectores soporte, seguido del de k vecinos próximos, y por último el de árbol de decisión. Sin embargo estos dos últimos obtienen resultados muy parejos y bastante cercanos al de SVM.

Para el modelo con extracción de características por medio de redes neuronales convolucionales, los resultados para los clasificadores basados en K-NN y SVM son muy parejos, quedando el basado en Árboles de Decisión con los peores resultados dentro de su modelo de extracción de estadísticas y de todos los clasificadores empleados en total.

Se sabe que los resultados pueden diferir en caso de emplear distintos parámetros de diseño a la hora de generar los distintos modelos de clasificación. La inclusión de una política de costes específica en el sistema puede bajar las prestaciones generales pero favorecer la de la detección de lesiones malignas, lo que contribuye a lograr el objetivo fijado en un principio.

Es cierto que uno de los objetivos planteados era la de generar un sistema capaz de localizar qué lesiones son benignas con el fin de evitar biopsias innecesarias, y los resultados no favorecen tanto este tipo de detecciones como la del melanoma, pero esta ha sido una decisión tomada a la hora de realizar la política de costes para favorecer la mayor detección de melanomas posible manteniendo el resto de resultados aceptables sin desviarnos de todos los objetivos fijados en el inicio.

El clasificador basado en Máquinas de Vectores Soporte es, como esperábamos, el más eficaz para realizar clasificación de datos referentes a imágenes, como es nuestro caso, tanto si utilizamos extracción de características diseñadas manualmente como si hacemos la extracción por medio de CNNs.

Es interesante comprobar que un procedimiento que requiere de la observación e interpretación de las imágenes bajo estudio, y como en este caso, de referencias médicas directas para la extracción de características como es el método tradicional, que ha de generarse además por medio de un proceso de desarrollo mucho más

tedioso, que el que emplea redes neuronales convolucionales, que obtiene resultados aunque inferiores, muy parejos a los del método tradicional. Esto da a entender por qué se están utilizando en muchas tecnologías que realizan clasificación técnicas de aprendizaje profundo, y se prefieren frente a modelos de aprendizaje automático tradicional.

Sin embargo, la combinación de ambos tipos de aprendizaje puede ser interesante en múltiples aplicaciones y el caso de nuestro sistema es una de ellas, pues muestra resultados parejos a los del procedimiento que se ha llevado a cabo hasta los últimos años, con unos tiempos de implementación mucho más reducidos, y que permiten obtener resultados óptimos para usuarios menos experimentados y cuyos conocimientos no son tan específicos acerca de los datos a tratar.

Es esperanzador además observar que podemos contar con diferentes métodos y tecnologías para lograr un fin determinado que contribuya a mejorar los sistemas y procedimientos clínicos.

Trabajar en proyectos que tienen posibilidades reales de ayudar en la lucha contra el cáncer, utilizando técnicas que hasta hace un tiempo no tenían tanta aplicación directa y que con el avance de nuevas tecnologías permitan invertir en la salud de los pacientes, es muy enriquecedor.

5.2.Líneas Futuras

Cuando se concluye un proyecto extenso, acerca del cual se ha meditado durante muchas horas, y sobre el cual se ha tomado una serie de decisiones en el proceso de desarrollo del mismo, se tiene en cuenta que existen otra serie de opciones que no se han tenido en cuenta en las distintas fases del trabajo, y que bajo los resultados obtenidos hacen plantearse si estás podrían contribuir a mejorar el sistema diseñado. A continuación se enuncian una serie de ideas que podrían desarrollarse con el fin de mejorar los resultados obtenidos o lograr nuevos objetivos.

En el caso de nuestro sistema, la información acerca de lo que se considera que son lesiones benignas o malignas se extraen a partir de las etiquetas de las mismas de las imágenes que las apoyan, ya que el objetivo era desarrollar un sistema destinado a trabajar en situaciones en las que se dispone de imágenes dermatoscópicas pero no necesariamente de más datos. Sin embargo la base de datos disponible, contenía una extensa lista de datos médicos referentes a las lesiones, y que podrían ser útiles para contrastar con la información contenida en las imágenes y favorecer a una mejora del sistema. El requisito de esto es precisamente que se necesitarían datos médicos en las futuras consultas realizadas con el sistema.

Sería interesante emplear, en el modelo de extracción de características diferenciadas manualmente, un análisis por componentes principales (PCA) que permitiera reducir el volumen de datos ofreciendo prestaciones similares o incluso superiores a las obtenidas.

Otra opción en caso de contar con un mayor volumen de la base de datos comprobar resultados con todos los descriptores diseñados manualmente con el objetivo de determinar la robustez que ofrece todo el conjunto de descriptores.

En caso de contar con un mayor volumen de datos y medios de procesamiento, sería interesante adaptar nuestros objetivos y sistema a modelos que emplearan aprendizaje profundo para generar sistemas completos de clasificación con el fin de mejorar las prestaciones disponibles hasta ahora y utilizar redes neuronales completas y no únicamente destinadas a la extracción de características como es nuestro caso.

Puede ser interesante emplear validación cruzada para no preocuparnos en la eficacia de los conjuntos de datos que diseñamos para hacer las pruebas y experimentos.

6. Marco Regulatorio

6.1. Análisis de la legislación aplicable sobre la implementación descrita en el trabajo

Haciendo referencia a las responsabilidades profesionales y éticas vertidas en el proyecto y según la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal, atendiendo al Título⁷ 2, y lo referente a los artículos que mencionamos a continuación realizamos la justificación del uso regulado de la base de datos médicos empleada en el desarrollo del Sistema de Ayuda al Diagnóstico Precoz de Melanomas.

En lo referente a los artículos:

- 6. Consentimiento del afectado.
- 8. Datos relativos a la salud.
- 11. Comunicación de datos.

El departamento de Teoría de la Señal de la Universidad Carlos III cumple los requisitos regulatorios referidos a la cesión y adquisición de los datos médicos utilizados y son cedidos al desarrollador del proyecto bajo el consentimiento del supervisor del mismo Fernando Díaz de María.

En lo referente al artículo 12: Acceso a los datos por cuenta de terceros; el desarrollador se compromete a hacer uso exclusivo de los datos cedidos por parte del departamento de Teoría de la Señal de la Universidad Carlos III de Madrid obtenidos para el desarrollo del proyecto, y de la eliminación de los mismos en el momento de la culminación del Trabajo de Fin de Grado.

⁷ <https://www.boe.es/buscar/act.php?id=BOE-A-1999-23750&p=20110305&tn=1#a6>

7. Entorno Socio Económico

7.1.Presupuesto de la elaboración del TFG.

A continuación realizaremos una estimación del coste total del proyecto. Se realizará un listado de los distintos recursos utilizados para el desarrollo del Sistema de Ayuda al Diagnóstico Precoz de Melanomas junto con sus costes asociados.

Recursos Físicos: Como recursos físicos únicamente disponemos de un PC portátil Sony Vaio Serie E SVE151 como herramienta de trabajo. El coste asociado del mismo aparece reflejado en la siguiente tabla:

Ordenador Portátil Sony Vaio Serie E SVE151	879 €
Total IVA incluido	879 €

Tabla 14: Presupuesto de Recursos Físicos.

Recursos Software: Los recursos software utilizados son: El sistema operativo Microsoft Windows 10, programas de ofimática de Microsoft Office 2010, software matemático Matlab R2017a. El coste asociado a los mismos se adjunta en la siguiente tabla:

Microsoft Windows 10 Home 64 bits	46,99 €
Microsoft Office Estudiantes 2010	159 €
Licencia Matlab Estudiantes	69 €
Total IVA incluido	274,99 €

Tabla 15: Presupuesto de Recursos software

Recursos Humanos: En lo referente a recursos humanos se considera que el desarrollador ha realizado el proyecto con nivel profesional equivalente a Ingeniero Junior con un coste de 15 € la hora, para el supervisor se considera un nivel equivalente a Ingeniero Senior con un coste de 20 € la hora.

El periodo de desarrollo del proyecto consta de aproximadamente 28 semanas en las que se ha trabajado de manera continua con una dedicación de 18 horas semanales por parte del desarrollador en trabajo directo sobre el proyecto y 2 horas semanales por parte del supervisor en tareas de control del mismo. A continuación se adjunta las

cantidades mencionadas ponderadas en un coste total asociado a los recursos humanos del proyecto completo.

	Horas/Semana	Coste/Hora	Coste Total
Desarrollador	18h/Semana	15€/h	7.560 €
Supervisor	2h/ Semana	20€/h	840 €
Total IVA incluido			8.400 €

Tabla 16: Presupuesto de Recursos Humanos

Total: Realizamos un cómputo final del coste total del proyecto completo a través de los presupuestos destinados en los recursos comentados anteriormente. El coste total se adjunta en la tabla siguiente:

Recursos Físicos	879 €
Recursos Software	274,99 €
Recursos Humanos	8.400 €
Total IVA incluido	9.553,99 €

Tabla 17: Presupuesto Total del Sistema se Ayuda al Diagnóstico Precoz de Melanomas

8. Referencias

- [E1]: <https://www.cancer.org/es/cancer/cancer-de-piel-tipo-melanoma.html>
.Sección del Melanoma en web de Organización Americana Contra el Cáncer, muestra múltiples referencias empleadas.
- [E2]: https://medlineplus.gov/spanish/ency/esp_imagepages/8912.htm . Aspectos generales de la Piel
- [E3]: <https://es.wikipedia.org/wiki/Piel> . Aspectos generales de la Piel
- [E4]: <http://www.lasexta.com/constantes-vitales/noticias/tasa-mortalidad-melanoma-estabiliza-espana-2016041800264.html> Estadísticas del cáncer en España.
- [E5]: <https://www.cancer.org/es/cancer/cancer-de-piel/prevencion-y-deteccion-temprana/que-el-cancer-de-piel.html> . Comentarios acerca de los distintos tipos de cáncer de piel.
- [E6]: <https://www.cancer.org/es/cancer/aspectos-basicos-sobre-el-cancer/que-es-el-cancer.html> Aspectos básicos del cáncer.
- [E7]: <https://www.cancer.org/es/cancer/cancer-de-piel-tipo-melanoma/acerca/que-es-melanoma.html> . Aspectos generales del Melanoma
- [E8]: <http://www.ideidermatologia.com/cancer-de-piel/prevencion/dermatoscopia-digital.html> .Dermatoscopia digital
- [E9]: <http://audioconsejosmedicos.blogspot.com.es/2010/05/manchas-en-la-piel-los-7-puntos-de.html> .Criterios de Glasgow
- [E10]: <http://www.euromelanoma.org/spain/examen-para-determinar-si-se-ha-desarrollado-c%C3%A1ncer-de-piel/%C2%BFen-qu%C3%A9-hay-que-fijarse>
.Criterios de detección de melanoma
- [E11]: <https://www.cancer.org/es/cancer/cancer-de-piel-tipo-melanoma/deteccion-diagnostico-clasificacion-por-etapas/como-se-diagnostica.html> . Etapas del Melanoma
- [E12]: http://www.dermamedicalsystems.com/index.php?menu_id=1 . Aparato Dermatoscópico MoleMAxHD
- [E13]: <http://criscancer.org/es/sobre-el-cancer/tipos-de-cancer/canceres-de-piel/?gclid=CJrVt66NrNUCFW4W0wodLKcHLQ> . Descripción técnica melanoma

- [E14]: <http://www.fotofinder-systems.com/> . Ejemplo de aparatos dermatoscópicos
- [E15]: <https://es.mathworks.com/discovery/deep-learning.html> .Deep learning
- [E16]: <http://www.infosalus.com/asistencia/noticia-inteligencia-artificial-puede-ayudar-mejorar-deteccion-precoz-cancer-piel-20170823173546.html> .
Tecnología que emplea aprendizaje automático para detección de Melanoma
- [E17]: <http://www.abc.es/salud/noticias/20140428/abci-cancer-piel-201404281403.html> . App para autoexamen de la piel, fotoskin
- [E18]: <http://www.montevideo.com.uy/contenido/Un-sistema-japones-de-IA-permite-el-diagnostico-temprano-del-cancer-de-piel-349856> . Tecnologías que emplean Deep learning en Análisis de Imágenes Dermatoscópicas.
- [E19]: <https://blog.statsbot.co/support-vector-machines-tutorial-c1618e635e93> . Acerca de SVM
- [E20]: <https://docs.gimp.org/es/glossary.html> .Componente H
- [E21]: https://www.nebrija.es/~cmalagon/inco/apuntes_mios/arboles_de_decision.pdf
Árbol de decisión

Anexo 1: English Summary

Chapter 1

Melanoma has some characteristics that make it different from other skin cancers. This kind of cancer is developed from melanocytes, cells that we can find also in normal skin in the moles.

If we diagnosed melanoma in the first phases of the illness, we would be able to remove it with minor surgery and we would avoid major surgery, which can put the patient's life at risk. On the other hand, in advanced phases of cancer, this kind of tumour can make connections through the blood to form metastasis, spreading cancer cells all over the body, putting the patient's life at risk.

Dermatologists often use survival measures to assess the prognosis of the cancer in each person. Analysing these measures, we can determine that the early diagnosis of a cancer would provide the salvation to the patient, but if we don't, cancer will spread to the lymphatic system, forming metastasis and the possibilities of recuperation will be significantly reduced. That is the reason for the importance of early diagnosis of cancer.

For all these reasons, it will be interesting to develop a system to help us detect this kind of cancer early, making it possible to easily reduce the risks associated with melanomas.

The specialist biopsies the lesion when melanoma is suspected. The biopsy produces a scar in patients' skin. It would be interesting to have a system to avoid unnecessary biopsies. Specifically, we mean a classification system which could use dermoscopy images of skin lesions to determine if the lesion is benign or not.

It could also serve as a teaching support for dermatologists with interest in digital dermoscopy, and could help physicians in primary care who are going to refer the patient to the specialist later.

The objective is to create a computer-aided diagnosis system for *in-situ* detection of melanomas in early phases.

It will serve as a support tool for the specialist who has to take a decision about the necessity of a biopsy in a suspicious lesion. The system will provide a second opinion to the specialist about the potential development of cancer in a lesion.

We do not have an intention to develop a definitive system of diagnostic with an absolute reliability, but something to help the physicians to make better decisions.

Chapter 2

Melanoma develops into the skin. Skin is formed by three layers: epidermis, dermis and hypodermis.

Epidermis is in turn divided into some layers: cornified layer, clear/translucent layer, granular, spinous, and basal/germinal layer.

Melanocytes are the cells responsible for generating melanin, a pigment that gives colour to the skin. These cells are localized at stratum basal, interleaved between basal cells.

The amount and distribution of melanin pigment in the epidermis is the main reason for variation in skin colour. Melanin is found in the melanosomes, particles formed in melanocytes. The number of melanosomes varies between racial groups: white people have less than black people.

When skin is exposed to sun rays, melanocytes generate more melanin into melanosomes and skin becomes darker. This is the mechanism used by melanin to stop sun rays, protecting our skin.

The more important skin cancers are basal cell cancer, squamous cell cancer and melanoma. Of all types of cancer, melanoma is the least frequently diagnosed but the most aggressive.

Melanoma is able to develop all over the skin, but is more frequently found at the areas most exposed to the sun. Ultraviolet radiation from sun rays can damage the DNA of skin cells, changing some genes that allows melanoma to develop.

If we diagnose this special cancer in the first phases of the illness, it will be easy to remove it. However, if we diagnose it in more advanced phases of the illness, metastasis could be formed and that situation put at risk the life of the patient.

Risk factors to develop a melanoma: Family history or personal background of melanoma or another kind of cancer, ultraviolet ray exposure without control, white skin, to have moles or freckles, to be blonde, any immunodeficiency, to be a man, old or xeroderma pigmentosum.

To favour early diagnosis, physicians recommend to exam our skin, to know it and be able to recognize changes in our moles. If we have one or more of risk factors to develop skin cancer, we should be controlled by a dermatologist, making possible to diagnose the cancer in the first phases.

The most used criterion by the specialists to detect the melanoma is known as ABCD rule, whose name stands for Asymmetry, Irregular Edge, Colour, Diameter, and Evolution of the lesion. It represents the aspects of the lesion which the specialist has to control to detect the melanoma.

Another used criterion is the known as 7 Glasgow points, which in turn can be divided into two types of criteria, major and minor criteria. The major criteria are: colour changes, size changes, and form changes. The minor criteria are diameter bigger than 7mm, bled inflamated or sensibility changes. If one major criterion is met, two points are added; if one minor criterion is met, one point is added. Then, if the lesion score reaches 3 points or more is suspicious.

The melanoma can be classified into 4 phases: phase 1, cancer with thickness less than 2 mm, phase 2, cancer with thickness more than 2 mm, phase 3, cancer with transmission to lymph node and phase 4, cancer with metastasis.

Nowadays, the specialists use the digital dermatoscope for lesion analysis. This tool allows to take high quality photos from the lesions, save them, analyse them. Today portable digital dermatoscopes exist as well as specific technologies for lesion analysis.

Moreover, promising technologies are being developed for early detection of melanoma. An example of this technologies are mobile apps which can ease the auto exam that patients do, and machine learning and, in particular, deep learning.

Chapter 3

The proposed computer-aided diagnosis system for early melanoma detection is a system which relies on image processing, machine learning and deep learning. The system can generate a model which classifies the images of a data set either as benign lesions or melanomas.

The system has been designed based on an image data base of dermoscopy images labelled by specialists.

We have proposed two models for feature extraction (which are compared), one of them manually designed and the other model which uses convolutional neural networks.

The extracted features are used as inputs to a binary classifier. In particular, we have considered three models of classification, namely: Decision Tree, K-NN and Support Vector Machine (SVM).

The model which uses manually designed features has a pre-processing phase where the images are segmented. Furthermore, it consider three different type of descriptors: colour descriptors, form descriptors and texture descriptors.

The color descriptors are: mean and variance of RGB and HSV components, RGB percentiles, green region area, bluish-gray region area and RGB component median. The form descriptors are: Area, Perimeter, Eccentricity and Circularity. The texture descriptors are: Contrast, Homogeneity, Entropy, Correlation and Energy extracted from GLCM. In total, 38 descriptors have been designed and 35 have been finally used.

The model which uses features extracted by CNNs does not need pre-processing phase. The images are fed into a pre-trained convolutional neural network and the features are obtained as the output of an intermediate layer of the network.

The training process of the classification models and the design parameters of the classifier for two type of features extraction are similar.

Chapter 4

The database used consists of 724 images of benign lesions and melanomas. The total number of benign lesions is 514 and the number of melanomas is 210.

The data base also contains the images masks which are used for segmentation, and medical information about the lesion such as the label which identifies the lesion as benign or melanoma.

Since the database is not balanced we have generated reduced balanced versions of the training (314 images) and test sets (104 images).

The cost policy used in the classification process penalizes more the melanomas not detected as melanomas than benign lesions not detected like benign.

The best of the classifiers tested for manually designed features was the SVM. It achieved the following results: AUC: 0.85; precision 0.71; sensibility: 0.83, specificity 0.65, and F1-score 0.76.

The best of the classifiers tested for CNN-based features was the SVM again. It achieved the following results: AUC: 0.81; precision 0.68; sensibility: 0.79; specificity 0.63; and F1-score 0.73.

The errors made when classifying benign for both models appear in lesions which show blue tonalities, non-homogeneous textures and morphologic irregularities.

The errors made when classifying melanomas appear in lesions which show tonalities that seems like those of benign lesions, without blue tones and with homogeneous textures.

Chapter 5

The best results are obtained by the model that uses manually designed features.

The implementation of the system that uses the extraction of features by neural networks does not need a previous segmentation of the lesion and gets similar results.

The best results for each model are obtained by the classification system based on Support Vector Machines.

In the future, it could be interesting to take advantage of medical data available in the current database in order to make the system more robust. Furthermore, PCA could be used to make an automatic selection of characteristics and more powerful deep learning tools could be used.

Anexo 2: Referencias sobre el código utilizado

Obtención de imágenes a partir de la matriz de datos médicos

```
for i=1:100
z_imagenes(:,:,i)=imread(z_lesiones(i).ImPath);
end
```

Obtención de las máscaras de las imágenes a partir de la matriz de datos médicos:

```
for i=1:100
z_mascaras(i)=load(z_lesiones(i).MaskPath);
end
```

Normalización de las características obtenidas a partir de los descriptores:

```
for i=1:35
test_norm(i,:) = zscore(img_test(i,:));
train_norm(i,:) = zscore(img_train(i,:));
end
```

Varianza de las componentes RGB y HSV:

Por medio de un bucle y un bucle anidado que recorran cada uno de los píxeles de la imagen, únicamente teniendo en cuenta aquellos píxeles distintos de cero, realizamos las sumas que aparecen en la igualdad anterior, tanto la que eleva el valor del píxel en cuestión x_i al cuadrado como la que no. Al acabar el bucle obtenemos el valor de la varianza introduciendo los valores obtenidos en el bucle, junto μ^2 , μ y n .

Descriptores basados en Mediana de componentes RGB

Para cada una de las componentes R, G, y B de una imagen obtengo el vector que contiene su histograma por medio de `c_B=imhist(Componente)`, calculo la mediana sobre ese vector por medio de la función `Mediana_B=median(c_B)` y me devuelve el valor en escala de gris que representa la mediana de los píxeles del histograma.

Éste es otro de los descriptores propuestos bajo consideraciones del desarrollador.

Descriptores basados en regiones verdosas de la lesión

Para poder desarrollar este proceso únicamente obtenemos la componente H de la imagen bajo estudio, y por medio de un bucle y un bucle anidado recorreremos todas las posiciones de la matriz de componente H obtenida. Para aquellas posiciones cuyo valor se encuentre en el rango entre 59/360 y 169/360 un contador irá aumentando, representando al final de los bucles la cantidad de píxeles con la tonalidad deseada.

Descriptor basado en Área de la lesión

```
a_Area_ = regionprops(z_M_pruebas, 'Area');
```

El primer parámetro a introducir es la matriz sobre la que se quiere extraer una información acerca de una propiedad definida en la función, en este caso la máscara. El segundo parámetro, es el nombre de la propiedad que deseamos obtener para la matriz introducida, todos estos nombres o valores están predefinidos y descritos en cualquier documentación oficial acerca de *regionprops()* de Matlab, en este caso la propiedad deseada es el Área.

Para obtener este tipo de propiedad es relevante tener en cuenta que valor introducir como segundo parámetro en esta función ya que varias obtienen áreas de regiones, pero no todas son válidas en todos los casos. Sin embargo en imágenes binarias, estas consideraciones no son tan relevantes ya que por ejemplo 'Area' y 'FilledArea' obtienen el mismo valor.

Descriptor basado en Perímetro de la lesión

Para poder obtener el perímetro de la lesión utilicé la misma herramienta que en el caso del Área:

```
Perimetro_ = regionprops(z_M_pruebas, 'Perimeter');
```

El valor a introducir será 'Perimeter'

Elipsoidalidad de la lesión

Al igual que el área y el perímetro, la Elipsoidalidad o excentricidad de la lesión se obtiene por medio de *regionprops()*:

```
Elipsoidalidad_ = regionprops(z_M_pruebas, 'Eccentricity');
```

Matriz de coocurrencia

Para poder obtener la matriz de coocurrencia de las imágenes bajo estudio, se transforman las imágenes a escala de grises y posteriormente y se emplea la función predefinida en Matlab *graycomatrix()* para calcular la matriz de coocurrencia de la imagen. Posteriormente para calcular los estadísticos a partir de la misma se utiliza contenido descargable de Matlab, la función *GLCM_Features4(z_GLCM2)* a la cual se le introduce como parámetro la matriz de coocurrencia en cuestión y te devuelve una estructura con los diferentes estadísticos calculados referentes a esa matriz.

Por ello para obtener los descriptores referentes a la matriz de coocurrencia solo se tenía que acceder a los valores que interesaban de la estructura devuelta por la función anteriormente mencionada.

Contraste

El siguiente código es un ejemplo de cómo se obtiene el estadístico de contraste a partir de la matriz de coocurrencia *GLCM2* en Matlab.

```
stats = GLCM_Features4(GLCM2,0);  
Contraste=stats.contr;
```

Para el resto de descriptores el proceso sería similar.

Entropía

Se obtiene en Matlab según:

```
o_Entropia=z_stats.entro;  
Se asume que  $0 \cdot \log(0) = 0$ .
```

Clasificación por medio de árbol de decisión

La implementación en Matlab se realiza de la siguiente manera:

```
TREE = fitctree(img314',etiq_train);  
Y_arbol= predict(TREE,img104');
```

Se introducen las características, `img314'`, y las etiquetas referentes a estas características, `etiq_train`, en el proceso de entrenamiento, se genera un modelo entrenado, `TREE`, que se introduce en la función `predict()` junto con las características de test `,img104'`, y realiza la clasificación entregando las etiquetas, `Y_arbol`, que considera según el modelo introducido.

K-NN

La implementación en Matlab se realiza de la siguiente manera:

```
W=fitcknn(img314',etiq_train,'NumNeighbors',5,'Distance','cosine','Cost',[0 1;2 0]);  
  
[Y_KNN,score,cost] = predict(W,img104');
```

Se introducen las características, `img314'`, y las etiquetas referentes a estas características, `etiq_train`, en el proceso de entrenamiento, se genera un modelo entrenado, `w`, que se introduce en la función `predict()` junto con las características de test `,img104'`, y realiza la clasificación entregando las etiquetas, `Y_KNN`, que considera según el modelo introducido.

SVM

La implementación en Matlab se realiza de la siguiente manera:

```
SVMModel=fitcsvm(img314',etiq_train,'KernelFunction','linear','Cost',[0 1;2 0]);  
  
Y_svm= predict(SVMModel,img104');
```

Se introducen las características, `img314'`, y las etiquetas referentes a estas características, `etiq_train`, en el proceso de entrenamiento, se genera un modelo entrenado, `SVMModel`, que se introduce en la función `predict()` junto con las características de test `,img104'`, y realiza la clasificación entregando las etiquetas, `Y_svm`, que considera según el modelo introducido.

